

ABSTRACT

Title of thesis: MODELING AND OPTIMAL CONTROL OF
 ESCHERICHIA COLI GENETIC CIRCUITS
 WITH INTRINSIC STOCHASTICITY

Jianfeng Du, Master of Science, 2006

Thesis directed by: Professor Evangelhos Zafiriou
 Department of Chemical and Biomolecular Engineering

This work considers aspects of the pathway performance optimization of *Escherichia coli* genetic circuits. A characteristic of such circuits is that some molecules are present in very low quantities. This leads to the use of a simulation environment, Stochastic Petri Nets (SPNs), which is appropriate for obtaining stochastic uncertainty information; however the direct use of such models in optimization poses significant computational problems. We propose to approximate SPNs by Langevin-type models that incorporate stochastic variance information generated by SPN simulations, but which are simpler and potentially applicable for direct use in optimization. We discuss such models in the context of σ^{32} -mediated stress responses. Simulation results for both ethanol stress response and recombinant protein production match well with experimental data from the literature. The optimal control problem that maximizes the production of active protein through antisense RNA induction is then studied.

MODELING AND OPTIMAL CONTROL OF *ESCHERICHIA COLI*
GENETIC CIRCUITS WITH INTRINSIC STOCHASTICITY

by

Jianfeng Du

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2006

Advisory Committee:

Professor Evangelos Zafiriou, Chair
Professor William E. Bentley
Professor Raymond A. Adomaitis

© Copyright by

Jianfeng Du

2006

TABLE OF CONTENTS

List of Tables	iii
List of Figures	iv
1 Introduction to Genetic Circuits in <i>Escherichia Coli</i>	1
1.1 About <i>Escherichia Coli</i>	1
1.2 σ^{32} Stress Circuit	1
1.3 Recombinant Protein Expression	6
1.4 Antisense RNA	8
2 Kinetic Models of Genetic Circuits	11
2.1 Introduction	11
2.2 Chemical Master Equations	13
2.3 Stochastic Petri Nets	16
2.4 Development of SPN Model for σ^{32} -mediated Ethanol Stress	18
2.5 Development of SPN Model for Recombinant Protein Production	26
3 Langevin-Type Models	32
3.1 Introduction	32
3.2 Proposed Linear Langevin Model Fit	34
3.3 Proposed Nonlinear Langevin Model Fit	37
3.4 Langevin Model for σ^{32} -mediated Ethanol Stress	41
4 Optimal Control of Genetic Circuits	46
4.1 Introduction	46
4.2 Optimization Objectives and Control Inputs	46
4.3 Optimal Control for ODE Models	47
4.4 Optimization for Recombinant Protein Production on A Deterministic Model and Evaluation Based on Stochastic Simulations	51
5 Conclusions and Future Work	59
Bibliography	62

LIST OF TABLES

2.1	Rate constants in σ^{32} stress circuits. k_i ($i = 1, \dots, 20$) are deterministic rate constants, s_i ($i = 1, \dots, 20$) are stochastic rate constant. . .	21
2.2	Components in σ^{32} ethanol stress circuit and their initial tokens . . .	22
2.3	Genes in σ^{32} ethanol stress circuit	22
2.4	Components in the σ^{32} genetic regulatory circuit for recombinant protein expression and their initial tokens	26
2.5	Genes in the σ^{32} genetic regulatory circuit for recombinant protein expression	28
3.1	Results for diagonal Q	44
3.2	Results when q_{67} is used.	45
3.3	Results when q_{67} is used and the error is pushed to σ_{67}	45

LIST OF FIGURES

1.1	Protein synthesis	3
1.2	σ^{32} regulatory pathway of heat shock and ethanol stress. Adopted from Srivastava et al.(2001)	6
1.3	σ^{32} regulatory pathway of recombinant organophosphorus hydrolase production	8
2.1	Stochastic Petri Net of the reactions in (2.4)	17
2.2	SPN of the σ^{32} genetic regulatory circuit for ethanol stress. Adopted from Srivastava et al.(2001)	19
2.3	Fold increase of total σ^{32} ($\sigma^{32} + E\sigma^{32} + \sigma^{32}:J$) in ethanol stress response. Solid line: response without antisense. Dashed line: response with antisense. Surrounding dotted lines: \pm standard deviation. Points with error bars: experimental data (Srivastava et al., 2000). Error bars represent standard error from duplicate experiments.	23
2.4	Fold increase of GroEL protein in ethanol stress response. Solid line: response without antisense. Dashed line: response with antisense. Surrounding dotted lines: \pm standard deviation. Points with error bars: experimental data (Srivastava et al., 2000). Error bars represent standard error from duplicate experiments.	24
2.5	Dynamics of state variables in ethanol stress response. Solid line: response without antisense. Dashed line: response with antisense.	25
2.6	SPN of the σ^{32} genetic regulatory circuit for recombinant protein expression	27
2.7	Normalized level of total OPH (OPH+OPH:J+active OPH) following induction of recombinant protein. Solid line: response without antisense. Dashed line: response with antisense. Surrounding dotted lines: \pm standard deviation. Points with error bars: experimental data (Srivastava et al., 2000). Error bars represent standard error from duplicate experiments.	29
2.8	OPH activity following induction of recombinant protein. Solid line: response without antisense. Dashed line: response with antisense. Surrounding dotted lines: \pm standard deviation. Points with error bars: experimental data (Srivastava et al., 2000). Error bars represent standard error from duplicate experiments.	30

2.9	Dynamics of state variables in recombinant protein expression. Solid line: response without antisense. Dashed line: response with antisense.	31
4.1	Piecewise control input for OPH production	52
4.2	Dynamics of state variables in recombinant protein expression with optimal piecewise input: σ^{32} translation rate	55
4.3	Probability distributions of components at 30 min in recombinant protein expression with optimal piecewise input	56
4.4	Solid line: The final active OPH level (from SPN simulations) as a function of t_1 , i.e. the time of antisense induction. Surrounding dotted lines: \pm standard deviation.	58

Chapter 1

Introduction to Genetic Circuits in *Escherichia Coli*

1.1 About *Escherichia Coli*

Escherichia coli, usually abbreviated to *E. coli*, is one of the main species of bacteria that live in the lower intestines of warm-blooded animals. They are necessary for the proper digestion of food and are part of the intestinal flora. Its presence in water or food is an indication of fecal contamination. *E. coli* is a gram-negative, non-spore-forming, rod-shaped bacteria that grows best at 37°C. It is a facultative anaerobe that ferments lactose to acid and gas within 24 hours of incubation. A great deal is known about the biochemistry and genetics of *E. coli*, and it continues to be an important tool for biological research. *E. coli* is capable to grow rapidly on high density in simple media without the addition of growth factors, and it is one of the most widely used microorganism in bioengineering.

1.2 σ^{32} Stress Circuit

There are two major steps in protein synthesis, transcription and translation (Fig. 1.1). DNA is the carrier of genetic information for all complex organisms. The structure of DNA is generally found as a helical duplex made up of two strands. In transcription, one strand of the DNA double helix, the sense strand, is used as

a template by the enzyme RNA polymerase (RNAP) to synthesize a messenger RNA (mRNA). For eukaryotic cells, their DNA is enclosed in the cell's nucleus, which is separated from the cytoplasm by a nuclear membrane. The transcription happens inside the nucleus, then mRNA migrates from nucleus to cytoplasm and the translation happens in the cytoplasm. For prokaryotic cells, the nucleus is absent and there is only a single loop of DNA. In translation, a ribosome moves from codon to codon along the mRNA. Amino acids are added one by one, translated into polypeptidic sequences dictated by DNA and represented by mRNA. Newly synthesized proteins often undergo a final conformation adaptation in conjunction with chaperon proteins.

The RNA polymerase binds to a type of regulatory protein called sigma factor. The complex is referred to as holoenzyme. Under the direction of sigma factor, RNA polymerase recognizes the promoter and then transcribes specific DNA segment into mRNA. There are a variety of sigma factors that RNA polymerase can combine with. Under normal growth conditions, RNAP binds to a major sigma factor σ^{70} . The RNAP: σ^{70} complex transcribe the DNA segments which are necessary for normal cellular functions.

Organisms are subject to a plethora of environmental and metabolic stress conditions. Under stress conditions, some cellular structures are damaged and some proteins are misfolded or unfolded. The accumulation of denatured proteins is detrimental to normal cellular functions and may threaten the life of the cell. Regulatory systems can detect the abnormality and initiate responses that increase the resistance of the cell to stress conditions and repair denatured proteins. The first stress

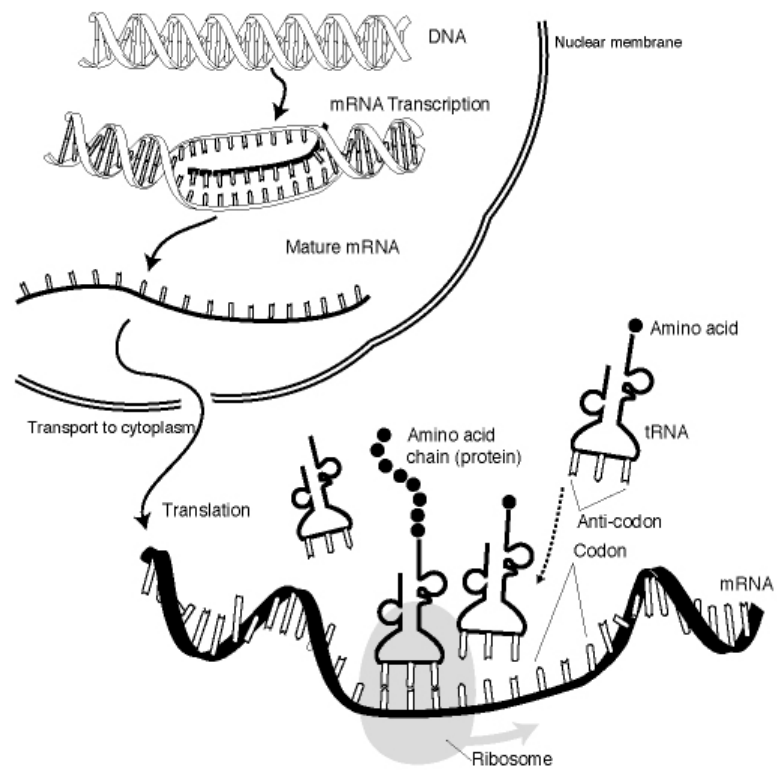


Figure 1.1: Protein synthesis

response mechanism observed was heat shock response (Neidardt, 1984; Neidardt, 1996).

When *E. coli* are exposed to high temperature, a set of proteins that fight against the stress are synthesized. These proteins have been collectively referred to as heat shock proteins, although a number of other stresses, such as chemical shock (ethanol, phenol, heavy metals), nutritional deficiency and induction of heterologous protein also initiate the production of these proteins in *E. coli* (Neidardt et al., 1996; Muffler et al., 1997; Gottesman, 1996). Under non-stress conditions, those heat shock proteins are present in *E. coli* at very low levels.

σ^{32} is an unstable sigma factor with a half life on the order of 1 minute (Strauss et al., 1987). The base level of σ^{32} is just a few molecules per cell in wild-type *E. coli* (Lesley et al., 1987). σ^{32} is encoded by the *ropH* gene in *E. coli*. Under normal growth conditions, σ^{32} is outcompeted by the major sigma factor, σ^{70} , thus σ^{32} mediated regulation does not play an important role. Under stress conditions, the level of σ^{32} is rapidly increased. Therefore, the σ^{32} mediated responses dominate the σ^{70} mediated responses.

The RNAP: σ^{32} complex, referred to as σ^{32} holoenzyme or $E\sigma^{32}$, recognizes the genes of heat shock proteins and transcribe them into mRNA. The σ^{32} mediated responses produce two types of heat shock proteins: chaperones and proteases. Chaperones help to refold the misfolded/unfolded proteins into their regular shapes. The chaperones that present in *E. coli* include GroEL, DnaK, DnaJ, GrpE, etc. Proteases target and destroy denatured proteins. The proteases produced in *E. coli* under stress conditions include FtsH, Lon, etc (Gottesman, 1996; Tomoyasu et al.,

1995; Blaszcak et al., 1995; Srivastava et al., 2001; Kurata et al., 2001; Gamer et al., 1992). Mutations that do not have *ropH* gene can not survive at temperatures above 20 °C (Zhou et al., 1988). The heat shock chaperones, DnaK, DnaJ and GrpE, combine with σ^{32} to form a complex. Because DnaJ- σ^{32} binding is rate limiting for complex formation, DnaK, DnaJ and GrpE are lumped as J protein in the following. The complex σ^{32} :J is presented to protease FtsH for degradation (Gamer et al., 1996; Strauss et al., 1989; Tomoyasu et al., 1995). While σ^{32} is destroyed, the J proteins are recycled after the reaction.

There are two reasons that there is a sharp increase of the σ^{32} level at the beginning of stress responses. First, the synthesis rate of σ^{32} increases immediately following the cells detect the stress conditions. Second, the sudden increase of denatured proteins compete for the J proteins that are required for the degradation of σ^{32} , resulting in an increased half life and stability of σ^{32} . (Strauss et al., 1989; Arsene et al., 2000; El-Samad et al., 2002). After a couple of minutes, the feedback mechanism begins to work and leads to the down regulation of the response. In this phase, the accumulation of J proteins and protease FtsH accelerate the degradation of σ^{32} , resulting in a decrease of σ^{32} level. Subsequently, the level of σ^{32} reaches a steady state because of the equilibrium between the increased translation of σ^{32} and the negative feedback by chaperones and proteases (El-Samad et al., 2002). The above mechanisms are illustrated in Fig. 1.2 and have been observed experimentally by Srivastava et al. (2001).

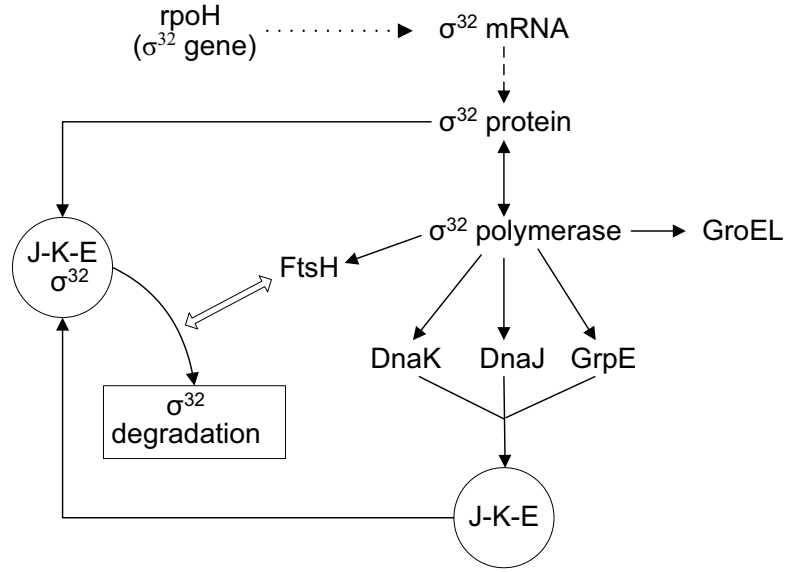


Figure 1.2: σ^{32} regulatory pathway of heat shock and ethanol stress. Adopted from Srivastava et al.(2001)

1.3 Recombinant Protein Expression

Genetic recombination refers to the exchange of genes between two DNA molecules to form new combinations of genes on a chromosome. Recombination of DNA occurs naturally in many microbes. Scientists developed artificial techniques to make recombinant DNA, which is an DNA sequence resulting from the combining of two other DNA sequences in a plasmid. A gene from humans can be inserted into the DNA of a bacterium, or a gene from a virus can inserted into a yeast. Recombinant proteins are proteins that are produced by genetically modified organisms following insertion of the foreign DNA into their genome. Some examples of recombinant DNA products are insulin, growth hormone, and oxytocin.

A plasmid is a circular double-stranded DNA molecules that replicate inde-

pendently from the chromosome. They usually occur in bacteria, sometimes in eukaryotic organisms and generally contain a genetic marker. Plasmids used in genetic engineering are called vectors, they can transfer genes from one organism to another. In a typical recombinant DNA procedure, a vector is isolated and a DNA is cleaved by restriction enzyme into fragments, then the desired gene is inserted into the vector *in vitro*. Next, the vector is taken by a cell such as a bacterium in order to express the gene and produce the protein coded for by the gene. Large amounts of the protein can be produced in a factory with vats of the genetically engineered bacteria. The recombinant DNA techniques can also be used to make thousands of copies of the same DNA molecule - to amplify DNA, thus generating sufficient DNA for various kind of experiment and analysis.

Recombinant proteins can be produced in many types of cells through genetic engineering. The commonly used cell types are bacterial, yeast, insect and mammalian cells (Srivastava, 1999). Among the many systems available for recombinant protein production, *Escherichia coli* is one of the most attractive because of its distinct advantages: it grows much faster than other cells because of its short life cycle; it can grow at high density on inexpensive substrate; its genetic and phenotypic properties are well characterized; a large number of cloning vectors and mutant host strains are available (Baneyx, 1999).

As mentioned earlier, the induction of heterologous protein triggers the σ^{32} mediated stress response. Compared with ethanol and heat shock responses, the production of recombinant protein induce the same set of heat shock proteins, suggesting there is some overlap among those regulatory pathways. The regulatory

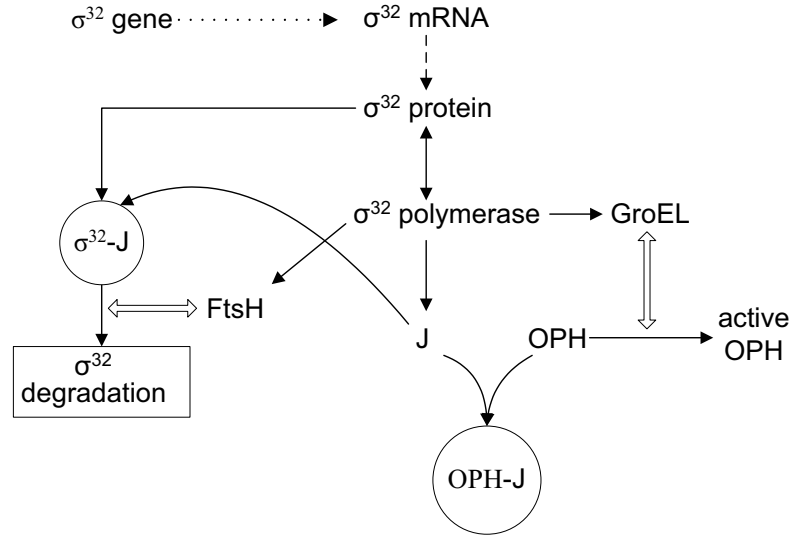


Figure 1.3: σ^{32} regulatory pathway of recombinant organophosphorus hydrolase production

pathway of organophosphorus hydrolase (OPH) production is illustrated in Fig. 1.3. After the recombinant protein is synthesized, it may compete with σ^{32} for the J protein, combine with J to form OPH:J complex. Some of the recombinant protein undergoes an activation process and becomes the active form. GroEL serves as the catalyst in the activation.

1.4 Antisense RNA

mRNA is transcribed from the sense strand of DNA and its structure is single stranded. Its sequence of nucleotides is called "sense" because it results in a gene product (protein). mRNA can form a duplex with a second strand of RNA whose sequence of bases is complementary to the first strand. The second strand is called antisense RNA. When mRNA forms a duplex with its complement, transla-

tion is down-regulated. This may occur because either the ribosome cannot access the nucleotides in the mRNA due to spatial obstacles or double-stranded RNA is quickly degraded by ribonucleases in the cell (Murray, 1992). Antisense RNA occurs naturally in some prokaryotic cells, it is also used artificially to downregulate gene expression. For example, antisense technology has been used to suppress the expression of certain enzyme in transgenic tomatoes, thus prolonging the ripening process of the tomato and its shelf life.

In the work of Srivastava et al. (2000), plasmids containing an antisense sequence of σ^{32} gene were constructed and introduced into *E. coli* cells. The effects of antisense RNA on σ^{32} mediated stress responses were evaluated.

Under ethanol stress, the antisense RNA was expressed by the introduction of plasmid pSE420 α s upon IPTG addition. Without antisense, there was a ten-fold increase of σ^{32} protein level after ethanol stress. However, in antisense producing cultures, a three-fold increase of σ^{32} was observed. Since GroEL is not involved in the metabolism of σ^{32} , it was used as a target protein to evaluate the regulation effects. Compared with no-antisense control cultures, GroEL levels were reduced significantly by antisense RNA, indicating downregulation of σ^{32} mediated stress response.

During the production of OPH, plasmid pTO, which expresses OPH only, and plasmid pTO α s, which coexpresses OPH and antisense RNA, were introduced into *E. coli*. The GroEL levels in antisense producing cultures were lower than no-antisense cultures, indicating downregulation of σ^{32} . The antisense RNA decreased the total yield of OPH to about two third of the yield without antisense RNA. However,

compared to no-antisense control cultures, a three fold higher OPH activity was observed with antisense.

Chapter 2

Kinetic Models of Genetic Circuits

2.1 Introduction

Gene expression is a complex process which is regulated at several stages in the synthesis of proteins. The regulation is achieved through multi-channel interactions between DNA, RNA, proteins and small molecules (Jong, 2002). To develop a deeper understanding for genetic regulatory systems, a variety of mathematical formalisms have been proposed by researchers.

The most simplified models of genetic circuits are Boolean networks. In Boolean networks, each gene is assumed to be either active (1) or inactive (0). For a genetic network with n genes, the size of state space is 2^n and generally can be listed exhaustly. The transitions between states are deterministic and synchronous, which can be expressed by a true table or a wiring diagram. Boolean networks' simplifying assumptions, (i.e. discrete time and state space, deterministic), limit their application in modeling genetic networks. In some situations, for example when transitions do not take place simultaneously or granular expression level information is necessary, Boolean networks are not appropriate (Gibson et al., 2001; Jong, 2002).

The kinetic behavior of chemical reactions is traditionally described by a set

of coupled ordinary differential equations (ODE):

$$\dot{x} = f(x) \quad x(0) = x_0 \quad (2.1)$$

where $x(t) \in R^N$ denotes the concentrations of the chemical components at time t , f is the phenomenological behavior of the system and is determined by stoichiometry of reactions, x_0 is the initial concentrations of the components. The implicit assumptions are: (i) the concentrations vary continuously, (ii) the reaction kinetics are deterministic. Such descriptions are valid when the amount of the reacting species is large and the thermal fluctuations do not contribute to the overall system behavior.

In genetic circuits, the intracellular concentrations of regulatory molecules are generally low and most individual genes are present in only one or two copies per cell. As a result, the discrete nature becomes conspicuous and stochastic effects may invalidate the deterministic equations. The stochastic nature of transcription has been experimentally observed in a number of cases (Goss and Peccoud, 1998). Modeling the gene expression mechanisms deterministically may be grossly misleading, especially if a kinetic trajectory lies close to a stable or unstable node of the system, in which case small fluctuations can be amplified and produce observable, even macroscopic effects. A more broadly applicable approach to model such systems is to use chemical master equations (CME), which describe the reaction network as a continuous time, discrete state-space and stochastic process. We will further discuss CME in the next section.

Depending on the specification of time, state-space and nature of determina-

tion, different kinetic models can be defined. Time can be discrete (D) or continuous (C), the state-space can be also discrete (D) or continuous (C), and the nature of determination can be deterministic (D) or stochastic (S) (Érdi and Tóth, 1988). Boolean networks can be identified as DDD models. Ordinary differential equations can be identified as CCD models. Chemical master equations can be identified as CDS models. DCD models have been widely used in modeling chaotic phenomena.

2.2 Chemical Master Equations

The chemical master equations can be used for stochastic simulation of chemical reaction systems that are well-stirred and in thermal equilibrium. For a system consisting of N species and M chemical reactions, the state of the system $x(t)$ is completely defined by the population of each species and is a jump Markov process. The CME describe the time evolution of a joint probability distribution $P(n, t)$, which is defined as

$$P(n, t) = \text{probability that } x(t) = n \quad (2.2)$$

The CME can be derived as (Gillespie, 1992)

$$\frac{\partial P(n, t)}{\partial t} = \sum_{j=1}^M [a_j(n - \nu_j)P(n - \nu_j, t) - a_j(n)P(n, t)] \quad (2.3)$$

where a_j is the propensity function of reaction j , ν_j is the state change vector of reaction j .

Let us consider a spatially homogeneous system composed of four chemical

species, A, B, C and D, and subject to two coupled chemical reactions:



where s_i ($i = 1, 2, 3$) are stochastic rate constants. They are related with deterministic rate constants k through:

$$s_1 = \frac{k_1}{VN_A} \quad (2.5)$$

$$s_2 = \frac{2k_2}{VN_A} \quad (2.6)$$

$$s_3 = k_3 \quad (2.7)$$

where V is the volume of the reaction system, N_A is the Avogadro's number. The propensity functions are

$$a_1(n) = s_1 n_A n_B \quad (2.8)$$

$$a_2(n) = s_2 \frac{n_B(n_B - 1)}{2} \quad (2.9)$$

$$a_3(n) = s_3 n_D \quad (2.10)$$

The state change vectors are

$$\nu_1 = (-1, -1, +2, 0) \quad (2.11)$$

$$\nu_2 = (0, -2, 0, +1) \quad (2.12)$$

$$\nu_3 = (0, +2, 0, -1) \quad (2.13)$$

We can write down the chemical master equations as

$$\begin{aligned}
\frac{\partial P(n_A, n_B, n_C, n_D; t)}{\partial t} = & \\
& s_1 [(n_A + 1)(n_B + 1)P(n_A + 1, n_B + 1, n_C - 2, n_D; t) - n_A n_B P(n_A, n_B, n_C, n_D; t)] \\
+ & s_2 \left[\frac{1}{2}(n_B + 2)(n_B + 1)P(n_A, n_B + 2, n_C, n_D - 1; t) \right. \\
& \left. - \frac{1}{2}n_B(n_B - 1)P(n_A, n_B, n_C, n_D; t) \right] \\
+ & s_3 [(n_D + 1)P(n_A, n_B - 2, n_C, n_D + 1; t) - n_D P(n_A, n_B, n_C, n_D; t)] \quad (2.14)
\end{aligned}$$

In principle, the solution of CME completely determines $P(x, t)$, and then $x(t)$. However, if the system involves more than a few molecular species and chemical reactions, the state space will be very large. For a three step reversible chain reaction



If there are 100 molecules involved in this reaction, the size of the state space is

$$C_{101}^3 + 2C_{101}^2 + C_{101}^1 = 176,851 \quad (2.16)$$

It means the CME consist of 176,851 differential equations. Thus CME are not tractable both analytically and numerically for all but the simplest systems.

Gillespie proposed two numerical methods using Monte Carlo techniques for the exact stochastic simulation of CME (Gillespie, 1976). In the Direct Method, the time of the next reaction τ is generated from the following exponential distribution

$$P(\tau) = ae^{-a\tau} \quad (0 \leq \tau < \infty) \quad (2.17)$$

where a is the overall propensity, $a = \sum_j a_j$. In determining which reaction will occur, the likelihood of a reaction being selected is proportional to its propensity

function

$$P(\mu) = \frac{a_\mu}{a} \quad (\mu = 1, \dots, M) \quad (2.18)$$

Then advance the system time to $t + \tau$, update the population of all species involved in reaction μ , and calculate the new propensity functions. This process is repeated until the “stopping time” of the simulation. Running the simulation multiple times gives the mean and variance information of state variables. The First Reaction Method generates a “tentative reaction time” τ_j for each reaction according to the following exponential distribution

$$P(\tau_j) = a_j e^{-a_j \tau_j} \quad (0 \leq \tau_j < \infty) \quad (2.19)$$

The reaction which occurs first is chosen as the “actual” next reaction, i.e.

$$\mu = j \quad \text{for which } \tau_j = \min(\tau_1, \dots, \tau_M) \quad (2.20)$$

Then the system states are updated as with the Direct Method and the process is repeated until the end of the simulation.

2.3 Stochastic Petri Nets

Stochastic Petri Nets (SPNs) were first introduced in the field of computer science and have been used to model biological systems (Goss and Peccoud, 1998; Srivastava et al., 2001). An SPN is composed of a set of places and transitions. Each place represents a chemical species and is drawn as a circle. Each place contains tokens, where the number of tokens is the number of molecules of the species. Each transition represents a chemical reaction and is drawn as a rectangle. The places and

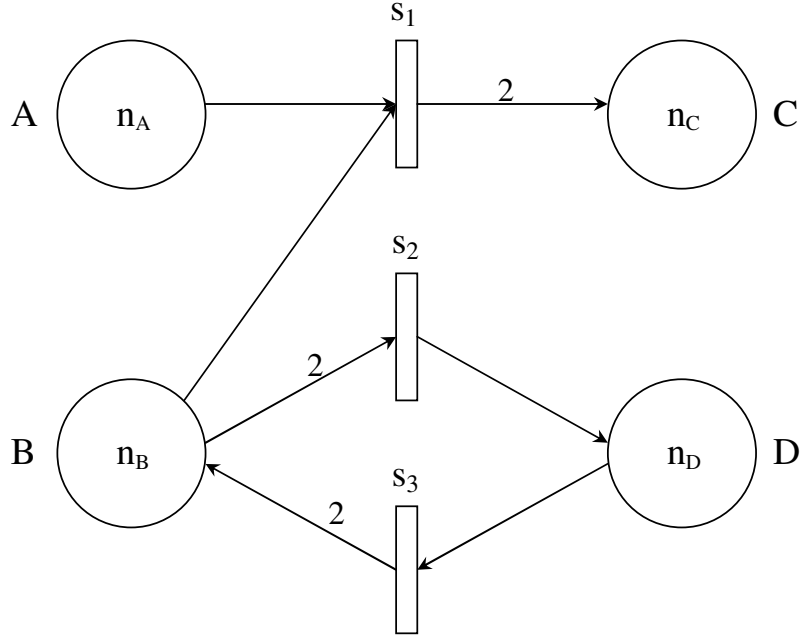


Figure 2.1: Stochastic Petri Net of the reactions in (2.4)

transitions are linked with arrows. If an arrow points from a place to a transition, the species represented by the place is a reactant. If an arrow points from a transition to a place, the species represented by the place is a product. If more than one token are consumed or produced in the transition, the arrow is labeled with the coefficient. The SPN of the reactions in (2.4) is shown in Fig. 2.1.

In SPN simulation, the transition j fires with an exponentially distributed time delay t_j

$$P(t_j) = w_j e^{-w_j t_j} \quad (0 \leq t_j < \infty) \quad (2.21)$$

The weight function w_j is determined from the stochastic rate constant and the number of tokens involved in the transition, and it is equivalent to the propensity function in CME. If more than one transition can fire, the transition with the smallest time delay is allowed to fire. Then the number of tokens involved in the transition

is appropriately updated, and weight functions are recalculated. The process is repeated until the end of the simulation. The implementation of SPN is equivalent to the First Reaction Method for the stochastic simulation of CME.

2.4 Development of SPN Model for σ^{32} -mediated Ethanol Stress

A Stochastic Petri Net (SPN) model was developed by Srivastava *et al.* (2001) for simulating the σ^{32} stress circuit. The model was validated against experiments in which ethanol (inducer of heat shock response) and σ^{32} -targeted antisense (downward regulator) were used to perturb the σ^{32} regulatory pathway. The regulatory pathway is shown in Fig. 1.2. The structure of the SPN is shown in Fig. 2.2. Except σ^{32} mRNA, the mRNA of other proteins are assumed to be constant numbers and are not explicitly represented. For those components, the transcription and translation are lumped into one reaction: synthesis.

In this work, the parameters of the SPN model are revised to eliminate some discrepancies in the original model and to further reduce the mismatch between simulation results and experimental data. One major change is the value of *E. coli* cell volume used in the calculation of stochastic rate constants for second order reactions. The average *E. coli* cell is a rod-shaped body with diameter and length of approximately 0.5 and 2 micrometers (Neidhardt, 1990). Thus the cell volume can be derived as $3.93 \times 10^{-19} \text{m}^3$. The GroEL synthesis rate, J synthesis rate and σ^{32} degradation rate are re-derived based on the new cell volume (Kanemori et al., 1994; Tomoyasu et al., 1995). The σ^{32} transcription rate and mRNA decay rate are

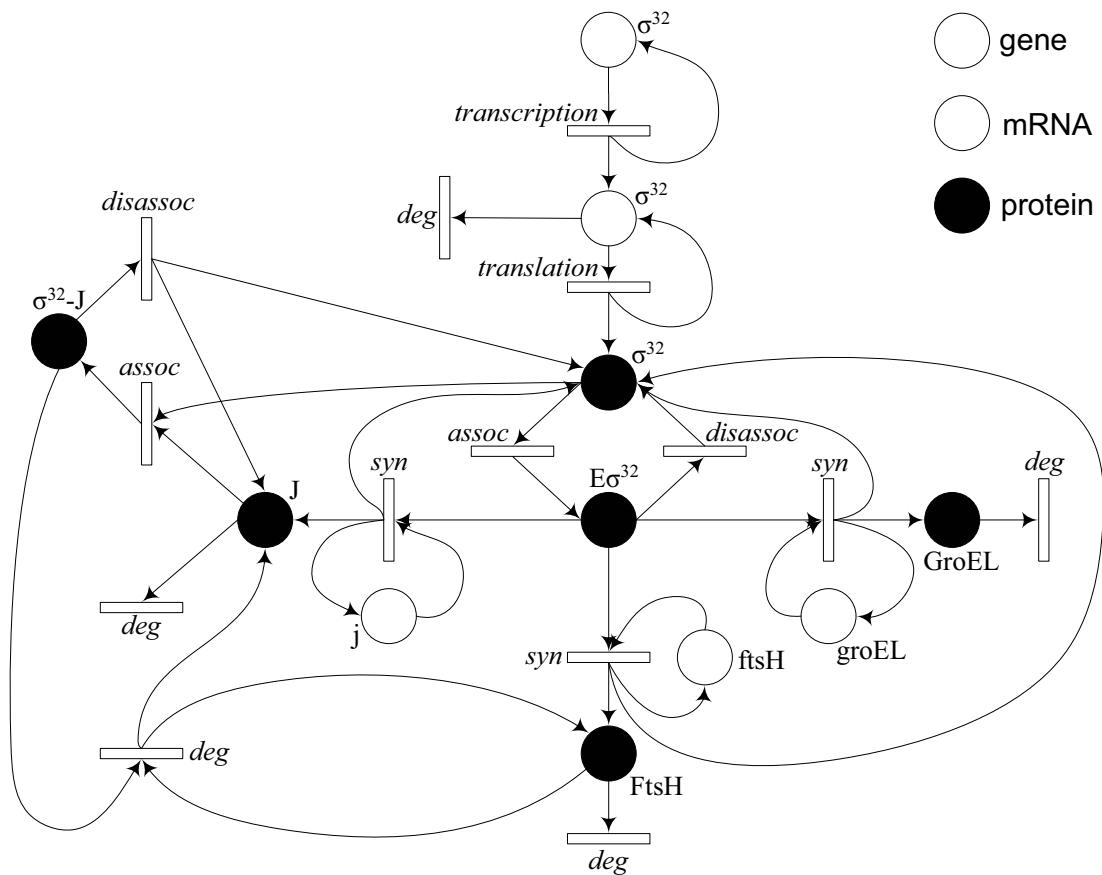


Figure 2.2: SPN of the σ^{32} genetic regulatory circuit for ethanol stress. Adopted from Srivastava et al.(2001)

adjusted to maintain σ^{32} mRNA at a constant level (10 molecules per cell). The σ^{32} translation rate is adjusted to $0.0051s^{-1}$ to get 9 σ^{32} molecules under nonstress conditions. The FtsH synthesis rate is adjusted to match the fold increase of total σ^{32} ($\sigma^{32}+E\sigma^{32}+\sigma^{32}:J$). To match the fold increase of GroEL protein, the number of initial GroEL tokens is adjusted to 1200. The deterministic and stochastic rate constants of the new model are listed in Table 2.1. The components in σ^{32} ethanol stress circuit and their initial tokens are listed in Table 2.2. All genes are present at one copy and keep constant during the stress response. They are listed in Table 2.3.

The SPN is simulated with the Möbius software package developed by the Performability Engineering Research Group (The Center for Reliable and High-Performance Computing at the University of Illinois, www.perform.csl.uiuc.edu). This is the successor software to the UltraSAN package (Sanders, 1995) developed by the same group and used in SPN simulations by Srivastava et al. (2001) and Goss and Peccoud (1998).

Experimentally it was found that antisense reached a maximum within 30 minutes postinduction and most of the metabolic activity occurred within this time frame (Srivastava et al., 2001). Thus the final time used in simulations was selected as 30 min. Also, the effects of cell division can be neglected in a 30 min timeframe. To simulate the effects of ethanol, the σ^{32} translation rate is raised from 0.0051, the normal growth rate, to 0.065. To simulate the antisense-mediated downregulation of the stress response, the translation rate is raised from 0.0051 to 0.01, resulting in a much weaker stress response. Compared to the experimental results (Srivastava et

Index	Transition	k	s
1	FtsH degradation	$7.40 \times 10^{-11} s^{-1}$	$7.40 \times 10^{-11} s^{-1}$
2	FtsH synthesis	$2.21 \times 10^8 (Ms)^{-1}$	$0.932 (Ts)^{-1}$
3	GroEL degradation	$1.80 \times 10^{-8} s^{-1}$	$1.80 \times 10^{-8} s^{-1}$
4	GroEL synthesis	$3.31 \times 10^9 (Ms)^{-1}$	$14.0 (Ts)^{-1}$
5	J-disassociation	$6.40 \times 10^{-10} s^{-1}$	$6.40 \times 10^{-10} s^{-1}$
6	J-production	$2.36 \times 10^9 (Ms)^{-1}$	$10.0 (Ts)^{-1}$
7	Holoenzyme association	$0.700 s^{-1}$	$0.700 s^{-1}$
8	Holoenzyme disassociation	$0.130 s^{-1}$	$0.130 s^{-1}$
9	σ^{32} mRNA decay	$1.40 \times 10^{-6} s^{-1}$	$1.40 \times 10^{-6} s^{-1}$
10	σ^{32} -J-association	$3.27 \times 10^5 (Ms)^{-1}$	$1.38 \times 10^{-3} (Ts)^{-1}$
11	σ^{32} degradation	$1.28 \times 10^4 (Ms)^{-1}$	$5.41 \times 10^{-5} (Ts)^{-1}$
12	σ^{32} -J-disassociation	$4.40 \times 10^{-4} s^{-1}$	$4.40 \times 10^{-4} s^{-1}$
13	σ^{32} transcription	$1.40 \times 10^{-5} s^{-1}$	$1.40 \times 10^{-5} s^{-1}$
14	σ^{32} translation	$5.10 \times 10^{-3} s^{-1}$	$5.10 \times 10^{-3} s^{-1}$
15	Recombinant protein-J-association	$7.11 \times 10^5 (Ms)^{-1}$	$3.01 \times 10^{-3} (Ts)^{-1}$
16	Recombinant protein-J-disassociation	$4.40 \times 10^{-5} s^{-1}$	$4.40 \times 10^{-5} s^{-1}$
17	Recombinant protein synthesis	$4.00 s^{-1}$	$4.00 s^{-1}$
18	Recombinant protein degradation	$3.00 \times 10^{-3} s^{-1}$	$3.00 \times 10^{-3} s^{-1}$
19	Recombinant protein activation	$3.27 \times 10^5 (Ms)^{-1}$	$1.38 \times 10^{-3} (Ts)^{-1}$
20	Active recombinant protein degradation	$3.00 \times 10^{-3} s^{-1}$	$3.00 \times 10^{-3} s^{-1}$

Table 2.1: Rate constants in σ^{32} stress circuits. k_i ($i = 1, \dots, 20$) are deterministic rate constants, s_i ($i = 1, \dots, 20$) are stochastic rate constant.

	Components	Initial tokens
x_1	σ^{32} mRNA	10
x_2	σ^{32}	1
x_3	$E\sigma^{32}$	1
x_4	GroEL	1200
x_5	FtsH	93
x_6	σ^{32} -J-Comp	7
x_7	J-Comp	54

Table 2.2: Components in σ^{32} ethanol stress circuit and their initial tokens

G_1	σ^{32} gene
G_2	GroEL gene
G_3	FtsH gene
G_4	J-Comp gene

Table 2.3: Genes in σ^{32} ethanol stress circuit

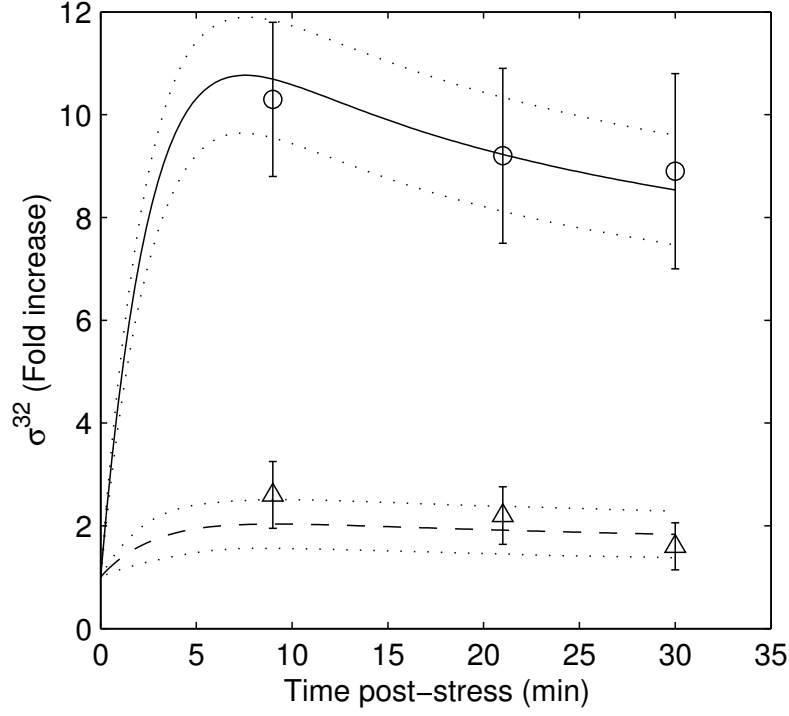


Figure 2.3: Fold increase of total σ^{32} ($\sigma^{32} + E\sigma^{32} + \sigma^{32}:J$) in ethanol stress response. Solid line: response without antisense. Dashed line: response with antisense. Surrounding dotted lines: \pm standard deviation. Points with error bars: experimental data (Srivastava et al., 2000). Error bars represent standard error from duplicate experiments.

al., 2000), the predictions of fold increases of σ^{32} and GroEL match the experimental data well (Fig. 2.3 and Fig. 2.4). The plots show the means and variances from SPN simulations. Compared with the means, the ODE simulation results are within 1-2% agreement. This agreement in general improves as more “batches” are used in Möbius simulations. The dynamics of state variables are shown in Fig. 2.5.

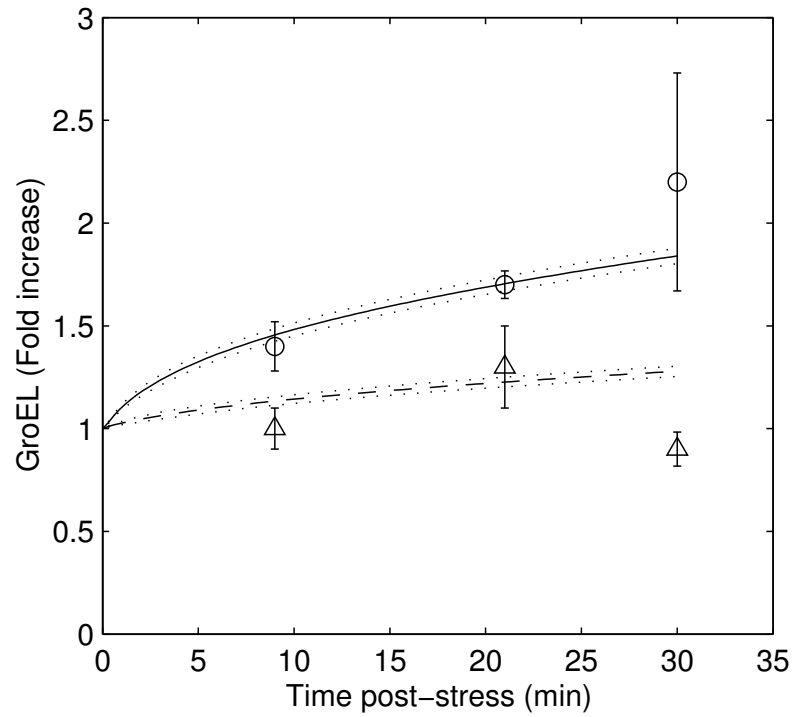


Figure 2.4: Fold increase of GroEL protein in ethanol stress response. Solid line: response without antisense. Dashed line: response with antisense. Surrounding dotted lines: \pm standard deviation. Points with error bars: experimental data (Srivastava et al., 2000). Error bars represent standard error from duplicate experiments.

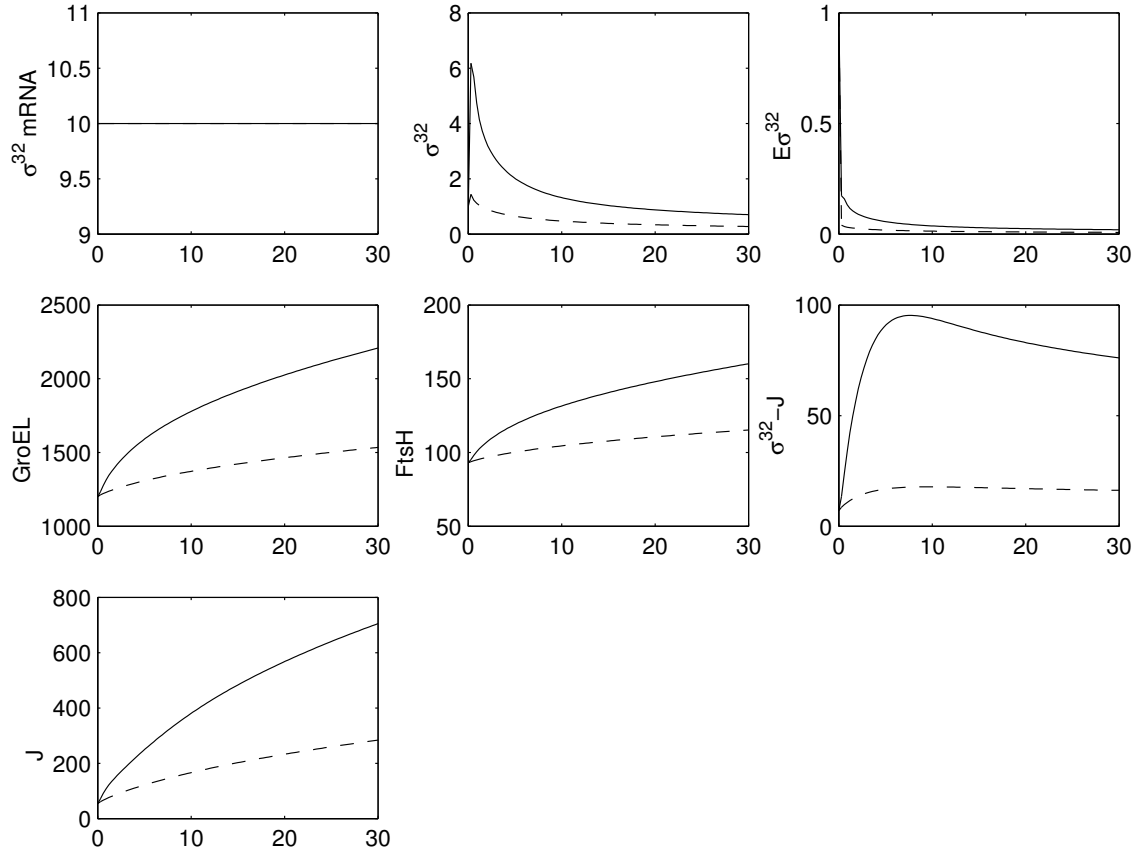


Figure 2.5: Dynamics of state variables in ethanol stress response. Solid line: response without antisense. Dashed line: response with antisense.

	Components	Initial tokens
x_1	σ^{32} mRNA	10
x_2	σ^{32}	1
x_3	E σ^{32}	1
x_4	GroEL	86
x_5	FtsH	93
x_6	Rec. Prot.	0
x_7	σ^{32} -J-Comp	7
x_8	J-Comp	54
x_9	Rec. Prot.-J-Comp	0
x_{10}	Active Rec. Prot.	0

Table 2.4: Components in the σ^{32} genetic regulatory circuit for recombinant protein expression and their initial tokens

2.5 Development of SPN Model for Recombinant Protein Production

The SPN model for recombinant protein production developed by Srivastava et al. (2001) is extended here to distinguish between active protein and inactive protein by adding one new component (active recombinant protein) and two new reactions (activation and degradation of the active recombinant protein) to the original SPN. The regulatory pathway is shown in Fig. 1.3. The new SPN is shown in Fig. 2.6. The components in the circuit and their initial tokens are listed in Table 2.4. The genes are listed in Table 2.5.

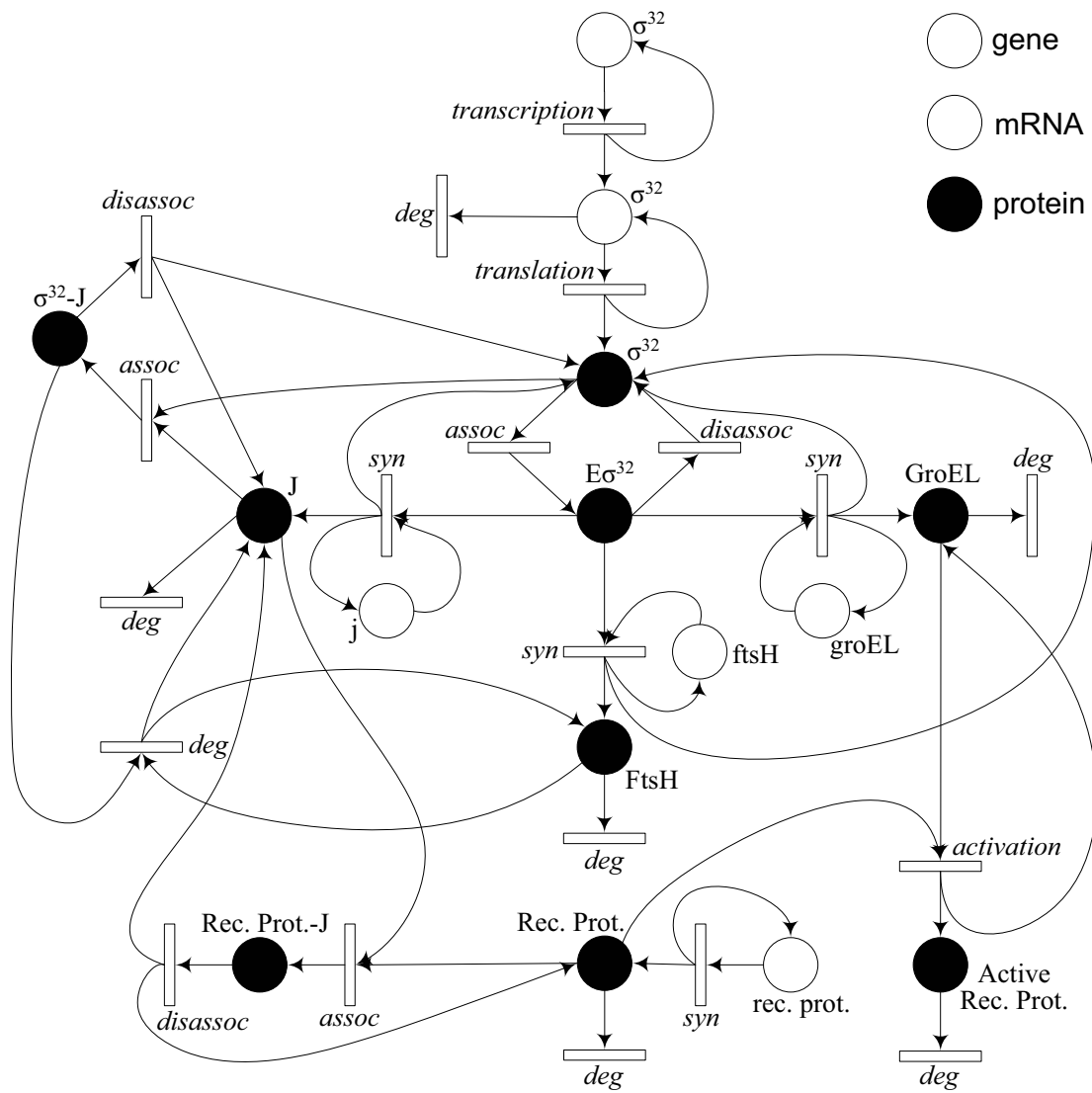


Figure 2.6: SPN of the σ^{32} genetic regulatory circuit for recombinant protein expression

G_1	σ^{32} gene
G_2	GroEL gene
G_3	FtsH gene
G_4	J-Comp gene
G_5	Rec. Prot. gene

Table 2.5: Genes in the σ^{32} genetic regulatory circuit for recombinant protein expression

The recombinant protein degradation rate is re-derived based on data from literature (Kanemori et al., 1994). The recombinant protein activation rate is assumed to be equal to the σ^{32} -J association rate. The active protein degradation rate is assumed to be equal to the inactive protein degradation rate. The GroEL synthesis rate has been adjusted downward to reflect the fact that its molecules form barrel-like groups of 14 molecules that essentially reduce the number of molecules that can be considered “active.” The number of initial GroEL tokens is reduced correspondingly. To simulate the effects of inducing production of recombinant protein with and without antisense, the σ^{32} translation rate is raised from 0.0051 to 0.14 and 0.01 correspondingly. Compared to the experimental results (Srivastava et al., 2000), the predictions of normalized level of total OPH (OPH+OPH:J+active OPH) and the protein activity match the experimental data well (Fig. 2.7 and Fig. 2.8). Note that in Fig. 2.8, the fraction of active OPH is matched to the activity measure for the experimental data by using a factor of 0.128 that provides the best fit with the data. The dynamics of state variables are shown in Fig. 2.9. We can

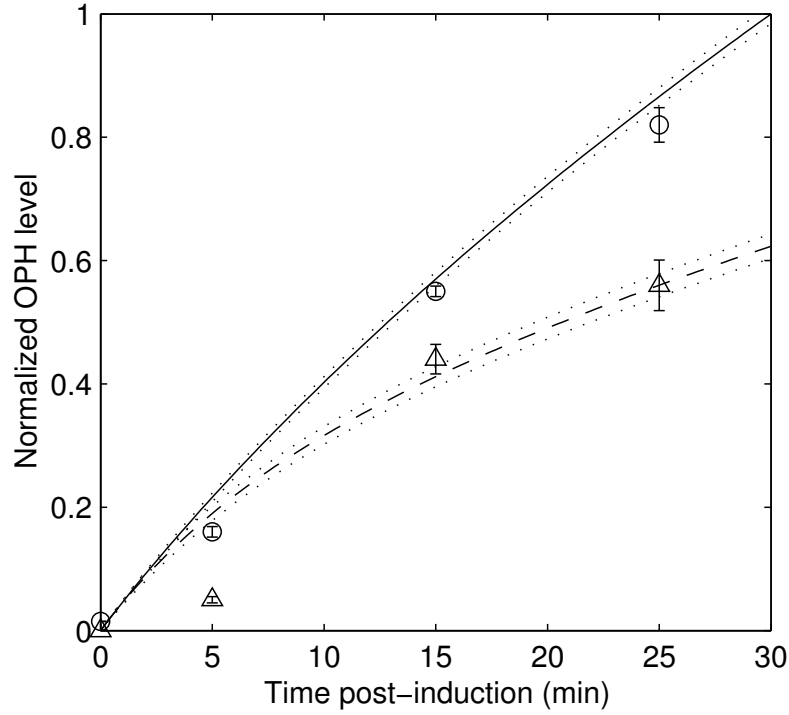


Figure 2.7: Normalized level of total OPH (OPH+OPH:J+active OPH) following induction of recombinant protein. Solid line: response without antisense. Dashed line: response with antisense. Surrounding dotted lines: \pm standard deviation. Points with error bars: experimental data (Srivastava et al., 2000). Error bars represent standard error from duplicate experiments.

see that the downregulation of σ^{32} -mediated response by antisense results in lower levels of J protein and total OPH. Since there is less J protein to combine with OPH, more OPH remains in free form, therefore more OPH is transformed into active form. This explains why there are both reduced total OPH level and increased OPH activity.

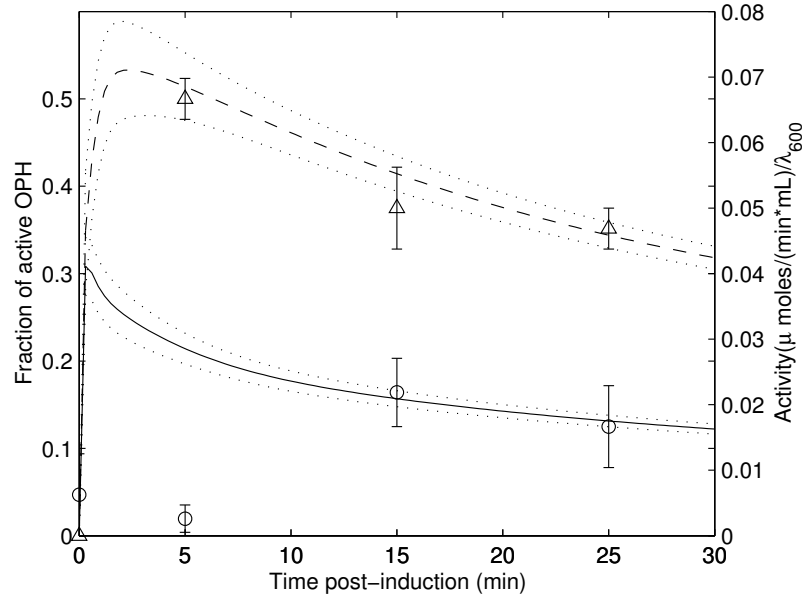


Figure 2.8: OPH activity following induction of recombinant protein. Solid line: response without antisense. Dashed line: response with antisense. Surrounding dotted lines: \pm standard deviation. Points with error bars: experimental data (Srivastava et al., 2000). Error bars represent standard error from duplicate experiments.

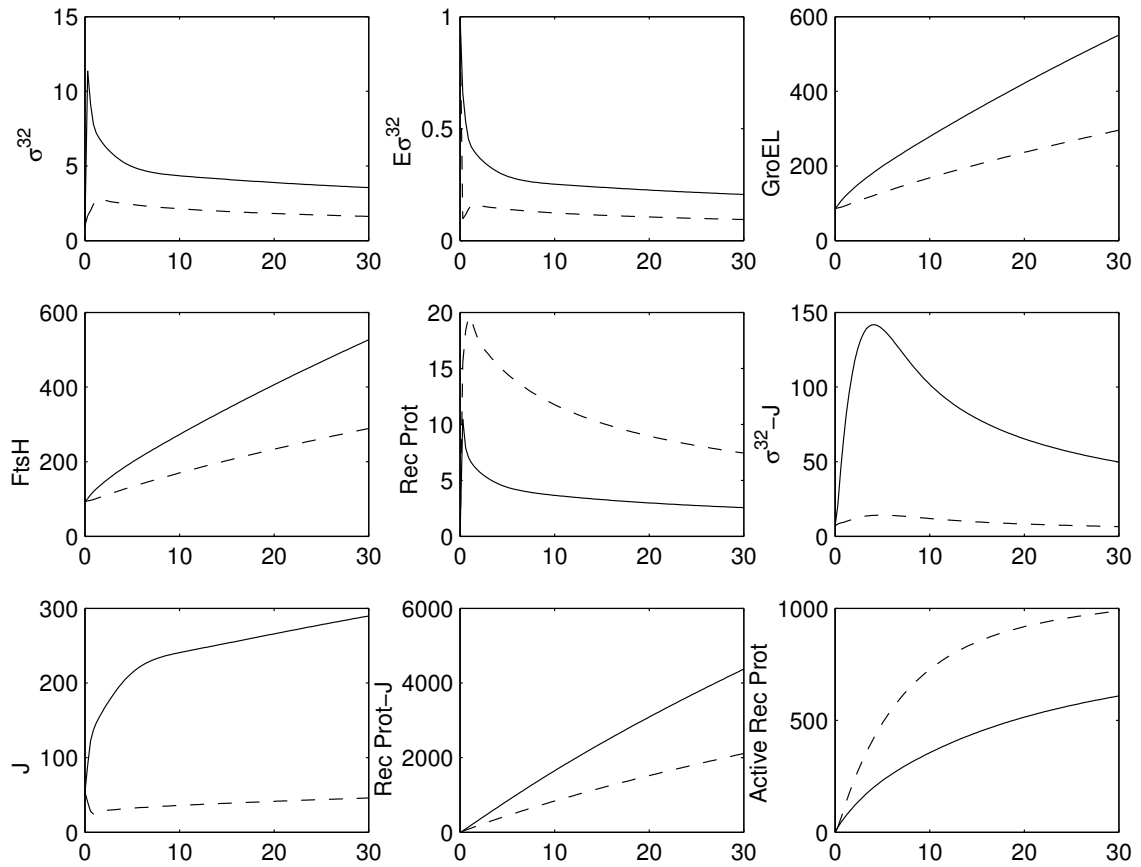


Figure 2.9: Dynamics of state variables in recombinant protein expression. Solid line: response without antisense. Dashed line: response with antisense.

Chapter 3

Langevin-Type Models

3.1 Introduction

SPN is an efficient method for modeling stochastic chemical reactions, and it has been successfully applied to several genetic systems. However, using an SPN model in an optimization framework is another matter. The representation of stochastic chemical reactions as a SPN is equivalent to their master equations, which describe the reaction network in continuous time and discrete state space. SPNs are isomorphic to hidden Markov chains. Including SPN as part of an optimization problem results in a Markov decision process (MDP). Solving an MDP problem is very difficult for all but the smallest problems. Indeed, Saucedo and Karim (1998) solved an optimal fermentor feed problem as an MDP, where the product concentration (ethanol) was discretized and the transition probability functions were obtained from an empirical input-output model. But this was a case with only one component present in the model. For the reaction networks of genetic circuits in *E. coli*, the state spaces generated by SPN simulation are extremely large. Solving the MDPs will be extremely difficult for these system, thus optimization based on an SPN model is impractical.

The Fokker-Planck equation was first introduced to describe the Brownian motion of small particles. The Fokker-Planck equation deals with fluctuations which

change the states of the system in an unpredictable but small way. This equation is now applied in a number of different fields such as solid-state physics, quantum optics, chemical physics, and circuit theory (Risken, 1989). The Fokker-Planck equation and the equivalent Langevin equation have been suggested as potentially useful for modeling at the gene level (Gibson and Mjolsness, 2001). In this work, Langevin-type models are proposed to approximate SPNs. Langevin models are suitable for use in optimization and sensitivity analysis, while at the same time they maintain the full set of reaction pathways as well as the stochastic nature of the system. The SPNs are used to generate the variance information which is necessary in developing Langevin models, as well as for testing theoretical predictions.

By allowing the number of molecules to take non-integer values, we can describe a reaction network by Langevin equation as

$$\dot{x} = f(x) + g(x)n(t) \quad (3.1)$$

where x is a vector of dimension N with each element corresponding to the number of molecules for each component present. The initial condition is that at time $t = 0$ the state variables have sharp values:

$$x(0) = x_0 \quad (3.2)$$

The function $f(x)$ is assumed to represent the phenomenological behavior of the system, i.e., the deterministic kinetics of the reaction network. The stochastic nature of the system is present in the Langevin equation via $n(t)$, which is a vector of noise variables with zero mean and with a correlation function which is proportional to a

δ function:

$$\begin{aligned}\langle n_i(t) \rangle &= 0 \quad i = 1, \dots, N \\ \langle n_i(t)n_j(t') \rangle &= q_{ij}\delta(t-t') \quad i, j = 1, \dots, N\end{aligned}\tag{3.3}$$

The symmetric matrix $Q = [q_{ij}]$ is referred to as the noise covariance matrix. The Langevin equation can be identified as continuous time, continuous state space, and stochastic model (CCS).

In this approach, g is assumed to be independent on x , therefore the constant g may be absorbed in Q and equation (3.1) can be simplified as

$$\dot{x} = f(x) + n(t)\tag{3.4}$$

(3.4) is called a Langevin equation with additive noises. By running SPN simulations, one can get the variance and covariance of x . Then the distribution information of x is used to determine the noise covariance matrix Q .

3.2 Proposed Linear Langevin Model Fit

For linear $f(x)$, the Langevin model is exactly equivalent to a continuous representation (non-integer molecule numbers) of the master equations that yield probability distribution. Take expectations of (3.4)

$$\langle \dot{x} \rangle = \langle f(x) \rangle + \langle n(t) \rangle\tag{3.5}$$

Since expectation is a linear operator and $\langle n(t) \rangle = 0$, we can get a deterministic differential equation with variable $\langle x \rangle$ from (3.5)

$$\dot{\langle x \rangle} = f(\langle x \rangle)\tag{3.6}$$

Thus the time evolution of the mean of x is exactly the solution of the phenomenological equations.

For linear $f(x)$, the Langevin equation (3.4) can be written as

$$\dot{x} = Ax + n(t) \quad (3.7)$$

where A is a constant matrix. A process described by (3.7) is called an Ornstein-Uhlenbeck process. When the dimension of x is one, (3.7) is the Langevin equation for Brownian motion.

The general solution of (3.7) is (Risken, 1989)

$$x(t) = \phi(t)x_0 + \int_0^t \phi(\tau)n(t-\tau)d\tau \quad (3.8)$$

where the $N \times N$ matrix $\phi(t)$ is the Green's function, which has to satisfy the matrix ODE

$$\dot{\phi} = A\phi \quad (3.9)$$

with initial condition

$$\phi(0) = I \quad (3.10)$$

The solution of the matrix ODE is

$$\phi(t) = I + At + \frac{1}{2}A^2t^2 + \dots \triangleq e^{At} \quad (3.11)$$

From (3.3) and (3.8), one can derive the equations of the distribution information of x (Risken, 1989):

$$M_i(t) = \langle x_i(t) \rangle = \sum_{j=1}^N \phi_{ij}(t)x_{0j} \quad i = 1, \dots, N \quad (3.12)$$

$$\begin{aligned}
\sigma_{ij}(t) &= \langle [x_i(t) - \langle x_i(t) \rangle][x_j(t) - \langle x_j(t) \rangle] \rangle \\
&= \sum_{k=1}^N \sum_{s=1}^N \int_0^t \int_0^t \phi_{ik}(\tau_1) \phi_{js}(\tau_2) q_{ks} \delta(\tau_1 - \tau_2) d\tau_1 d\tau_2 \\
&= \sum_{k=1}^N \sum_{s=1}^N \int_0^t \phi_{ik}(\tau) \phi_{js}(\tau) d\tau q_{ks} \quad i, j = 1, \dots, N
\end{aligned} \tag{3.13}$$

where $M_i(t)$ is the mean trajectory of x_i , $\sigma_{ij}(t)$ is the covariance of x_i and x_j . When $i = j$, the variance $\sigma_{ii}(t)$ can be generated from SPN simulation by Möbius. When $i \neq j$, there is no direct way to calculate the covariance of two variables by Möbius. However, covariance can be obtained indirectly. If we define a new performance variable which equals $x_i(t)x_j(t)$ and get its mean trajectory from Möbius, then the covariance of x_i and x_j can be calculated from the following equation:

$$\sigma_{ij}(t) = \langle x_i(t)x_j(t) \rangle - \langle x_i(t) \rangle \langle x_j(t) \rangle \tag{3.14}$$

Given the value of $\sigma_{ij}(t)$, (3.13) is a set of linear equations with N^2 variables and N^2 equations for any time point t , e.g. the final time of the simulation. The solution of (3.13) gives us the covariance information of the noises in the Langevin model (3.4), which should result in the same distribution for x as the results generated by an SPN simulation.

For the sake of explicit interpretation of noises in the context of robust control, we can further simplify the Langevin model by assuming a diagonal Q and just using the variance information of x to determine Q . Thus in (3.13) only N linear equations remain:

$$\sigma_{ii}(t) = \sum_{k=1}^N \int_0^t \phi_{ik}^2(\tau) d\tau q_{kk} \quad i = 1, \dots, N \tag{3.15}$$

The dimensions of variables and equations are still equal for a diagonal Q . This simplification may result in some negative q_{kk} in the solution of (3.15). Since the physical meaning of q_{kk} is the variance of noise n_k , it must be non-negative. If this happens, an alternative is to substitute the linear equations with a NLP problem subject to a bound on q , and the objective is to minimize the sum of squared relative errors

$$\min_q \sum_{i=1}^N \left[\frac{\sigma_{ii}(t) - \sum_{k=1}^N \int_0^t \phi_{ik}^2(\tau) d\tau q_{kk}}{\sigma_{ii}(t)} \right]^2$$

s.t. $q_{kk} \geq 0 \quad k = 1, \dots, N$ (3.16)

3.3 Proposed Nonlinear Langevin Model Fit

For nonlinear $f(x)$, (3.6) is not strictly valid since

$$\langle f(x) \rangle \neq f(\langle x \rangle) \tag{3.17}$$

Therefore difficulties arise in interpreting $\dot{x} = f(x)$ as the phenomenological (deterministic) equations of the reaction network. An example is the case of a diode circuit model where this interpretation leads to a violation of the second law of thermodynamics, i.e. Brillouin's paradox (Brillouin, 1950). The discrepancy is extremely small and of no practical interest (McFee, 1971), but the theoretical implications were Van Kampen's (1992) motivation for a method based on the expansion of the master equations. However, a Langevin model proposed here is not necessarily the means to a very accurate simulation. An SPN model can be used for that purpose

and to verify optimization results. Since Langevin models are just approximations to master equations, the interpretation difficulties are expected to not be important. A comparison of the time evolution of $\langle x \rangle$ in an SPN simulation to the path obtained by solving $\dot{x} = f(x)$ for the same initial conditions can be used to evaluate if this effect is important for the reaction networks under consideration. Actually, for both ethanol stress and recombinant protein production circuits the comparison showed excellent match. This is further supported by the simulation of ColE1 plasmid replication (Goss and Peccoud, 1998). This is a system where variance information is very important, so an SPN model was used to simulate it. The comparison showed excellent agreement between results obtained from the deterministic model (Brendel and Perelson, 1993) and the SPN-calculated mean for plasmid, RNA I and Rom protein. Of course the deterministic model alone could not provide the important variance information, but the agreement with the means obtained from the SPN simulation supports the interpretation of $\dot{x} = f(x)$ as the phenomenological equations for the system.

To determine the matrix Q for nonlinear reaction networks, we need to approximate the nonlinear equations in linear forms. Let us assume $x_m(t)$ is the mean trajectory that $x(t)$ follows, which can be obtained from an SPN simulation or by solving the phenomenological differential equations. Linearizing $f(x)$ in (3.4) around $x_m(t)$ results in a time-varying differential equation

$$\dot{x} = A(t)x + n(t) \tag{3.18}$$

where

$$A(t) = \left[\frac{df}{dx} \right]_{x=x_m(t)} \quad (3.19)$$

The general solution of (3.18) takes the same form as (3.8)

$$x(t) = \phi(t)x_0 + \int_0^t \phi(\tau)n(t-\tau)d\tau \quad (3.20)$$

where $\phi(t)$ has to satisfy the time varying matrix ODE

$$\dot{\phi} = A(t)\phi \quad (3.21)$$

with initial condition

$$\phi(0) = I \quad (3.22)$$

The matrix $\phi(t)$ is commonly referred to as the state transition matrix (STM) associated with $A(t)$. When $A(t)$ is independent of time, $\phi(t)$ is also called Green's function. Our approach for determining Q for linear Langevin models still applies for nonlinear $f(x)$ except for the calculation of $\phi(t)$. In general, it is not possible to derive an analytical, closed-form expression of the STM associated with an arbitrary matrix $A(t)$ (Tsakalis and Ioannou, 1993). The Peano-Baker formula is a power series expansion for the STM (Tsakalis and Ioannou, 1993):

$$\begin{aligned} \phi(t) = I &+ \int_0^t A(\tau_1)d\tau_1 + \int_0^t \int_0^{\tau_1} A(\tau_1)A(\tau_2)d\tau_2d\tau_1 + \dots \\ &+ \int_0^t \int_0^{\tau_1} \dots \int_0^{\tau_{n-1}} A(\tau_1)A(\tau_2) \dots A(\tau_n)d\tau_n \dots d\tau_2d\tau_1 + \dots \end{aligned} \quad (3.23)$$

However, there is no general rule to determine how many terms are sufficient for Peano-Baker formula to get a converged solution. Therefore, the matrix ODE (3.21) should be solved by numerical integral. The number of variables in (3.21) is N^2 , thus

the running of a ODE solver might be very slow. If this happens, we can decompose the matrix ODE into a set of smaller ODEs. Let us assume

$$\phi = [\phi_1, \dots, \phi_N] \quad (3.24)$$

where ϕ_i is the i th column of the matrix ϕ . Take derivatives of (3.24)

$$\dot{\phi} = [\dot{\phi}_1, \dots, \dot{\phi}_N] \quad (3.25)$$

From (3.21), (3.24) and (3.25), we obtain

$$\begin{aligned} \dot{\phi}_1 &= A(t)\phi_1 \\ &\vdots \\ \dot{\phi}_N &= A(t)\phi_N \end{aligned} \quad (3.26)$$

Thus the computation burden can be reduced greatly.

3.4 Langevin Model for σ^{32} -mediated Ethanol Stress

From the SPN of ethanol stress in Fig. 2.2, we can write down the Langevin model as a set of stochastic differential equations:

$$\begin{aligned}
\dot{x}_1 &= k_{13}G_1 - k_9x_1 + n_1(t) \\
\dot{x}_2 &= k_{14}x_1 + k_{12}x_6 - k_{10}x_2x_7 + k_6G_4x_3 - k_7x_2 + k_8x_3 + k_4G_2x_3 + k_2G_3x_3 + n_2(t) \\
\dot{x}_3 &= k_7x_2 - k_6G_4x_3 - k_2G_3x_3 - k_4G_2x_3 - k_8x_3 + n_3(t) \\
\dot{x}_4 &= k_4G_2x_3 - k_3x_4 + n_4(t) \\
\dot{x}_5 &= k_2G_3x_3 - k_1x_5 + n_5(t) \\
\dot{x}_6 &= k_{10}x_2x_7 - k_{12}x_6 - k_{11}x_5x_6 + n_6(t) \\
\dot{x}_7 &= k_6G_4x_3 + k_{12}x_6 - k_{10}x_2x_7 - k_5x_7 + k_{11}x_5x_6 + n_7(t)
\end{aligned} \tag{3.27}$$

where

$$\begin{aligned}
\langle n_i(t) \rangle &= 0 \quad i = 1, \dots, 7 \\
\langle n_i(t)n_j(t') \rangle &= q_{ij}\delta(t - t') \quad i, j = 1, \dots, 7
\end{aligned} \tag{3.28}$$

There are seven state variables in the Langevin model. They are listed in Table 2.2.

Let us not consider the cross variances and assume a diagonal Q . The variance for each state at the end of the SPN simulation time $t_f = 30\text{min}$ is obtained and related through the Langevin model to the variance of the noise terms in the model.

The result is a linear system of seven equations and seven unknowns:

$$Hq = \sigma \tag{3.29}$$

where $q = [q_{11}, \dots, q_{77}]^T$, is the vector of unknowns, $\sigma = [\sigma_{11}, \dots, \sigma_{77}]^T$ is a vector

with the state variances, and the matrix H is obtained from the Langevin model,

$$H_{ij} = \int_0^t \phi_{ij}^2(\tau) d\tau \quad (3.30)$$

Since the elements of q have to be non-negative, the solution is obtained as the result of the following optimization, where the scaling matrix V is a diagonal matrix composed of the elements of σ :

$$\begin{aligned} \min_q \quad & ||V^{-1}(Hq - \sigma)|| \\ \text{s.t.} \quad & q_{ii} \geq 0 \quad i = 1, \dots, 7 \end{aligned} \quad (3.31)$$

The result of the optimization is shown in Table 3.1. The match is good for most elements, except for σ_{33} and σ_{77} , where there are large errors. The errors are from the non-linearity of the system.

We should take note of the fact that the variances obtained from the SPN simulation also have related error in their calculation. The Möbius software package was set in the simulation to satisfy a 10% error constraint on the calculation with a desired 95% confidence. At the end of the simulation, we actually have bounds for the variances that Möbius produces and in principle we should only attempt to match within those limits. Furthermore, Möbius can be set up to indirectly provide cross-covariances for the states. Related error bounds can be computed afterwards and they tend to be larger than those for the state variances.

We have tried different alternative optimization formulations, which include the cross-covariance for states 6 and 7 ($\sigma_{67} = \sigma_{76}$) that correspond to components for which such a correlation may be significant. An additional variable in q is then

$q_{67} = q_{76}$, so we have in this case eight equations in eight variables

$$Hq = \sigma \quad (3.32)$$

Where $q = [q_{11}, \dots, q_{77}, q_{67}]^T$, $\sigma = [\sigma_{11}, \dots, \sigma_{77}, \sigma_{67}]^T$, H is a 8×8 matrix which has counted for the influence of q_{67} . Rather than trying to match exactly to the values in σ , these formulations attempt to satisfy the bounds obtained around these values from Möbius, and furthermore allow violation of these bounds, by setting them up in the optimization as soft constraints. Let us define $\bar{\sigma}$ to be upper bound of σ , $\underline{\sigma}$ to be the lower bound of σ , ϵ to be a vector of variables corresponding to violations, $\epsilon = [\epsilon_{11}, \dots, \epsilon_{77}, \epsilon_{67}]$. The objective function is to minimize the sum of square of violations, which are appropriately scaled by dividing with the values in σ , similarly to (3.31),

$$\begin{aligned} \min_q \quad & ||V^{-1}\epsilon|| \\ \text{s.t.} \quad & q_{ii} \geq 0 \quad i = 1, \dots, 7 \end{aligned} \quad (3.33)$$

or with the size of the constraint,

$$\begin{aligned} \min_q \quad & ||W^{-1}\epsilon|| \\ \text{s.t.} \quad & q_{ii} \geq 0 \quad i = 1, \dots, 7 \end{aligned} \quad (3.34)$$

or the square root of the product of the two,

$$\begin{aligned} \min_q \quad & ||(\sqrt{VW})^{-1}\epsilon|| \\ \text{s.t.} \quad & q_{ii} \geq 0 \quad i = 1, \dots, 7 \end{aligned} \quad (3.35)$$

where W is a diagonal matrix composed of the elements of $\bar{\sigma} - \underline{\sigma}$. These alternative formulations provided some improvement by reducing the error in matching diagonal

σ	q	$V^{-1}(Hq - \sigma)$
0.05	2.97×10^{-5}	0.0657
0.699	0	0.176
0.0203	0.156	-0.811
2040	0.518	6.57×10^{-5}
67.8	0.0363	0.0897
92.8	0.437	0.160
547	0	-0.648

Table 3.1: Results for diagonal Q .

elements. Table 3.2 shows the results when the square root of the product of the values in σ and the size of the constraint is used for scaling. Actually, if one allows pushing the error to the cross-covariance element σ_{67} , then the error in the match of the diagonal elements is small (Table 3.3). However, there is no general rule to determine how many non-diagonal elements to include and which ones to include. And the use of non-diagonal elements in Q may complicate the future use of the Langevin model in optimization.

σ	q	$V^{-1}(Hq - \sigma)$
0.05	3.34×10^{-5}	0.201
0.699	0	0.113
0.0203	0.0203	-0.0709
2040	2.97	0.00308
67.8	0.0491	-0.0830
92.8	1.22	0.141
547	4.77	-0.523
-97.2	-4.92	1.61

Table 3.2: Results when q_{67} is used.

σ	q	$V^{-1}(Hq - \sigma)$
0.05	3.51×10^{-5}	0.259
0.699	0	0.0655
0.0203	0.0197	-0.0993
2040	76.6	-0.00656
67.8	0.427	-0.0393
92.8	7.30	0.0632
547	12.8	-0.223
-97.2	-47.7	30.2

Table 3.3: Results when q_{67} is used and the error is pushed to σ_{67} .

Chapter 4

Optimal Control of Genetic Circuits

4.1 Introduction

After a Langevin model is obtained as an approximation of the master equations by using SPN-generated variance information, it can be used in an optimal control problem. The optimal control for genetic circuits in this work is to develop control policies for dynamically optimizing pathway performance by manipulating environmental factors and *in vivo* controllers. To deal with the stochastic nature of Langevin models, a robust control problem is generated by interpreting the noise $n(t)$ as “uncertainty” in the models. As the first attempt, optimal control policies are developed for the deterministic part of a Langevin model by assuming $n(t) = 0$. Then the effect of uncertainty on the optimality of these policies can be evaluated by maximizing the deviation of the value of the objective function from the deterministic optimal over all possible $n(t)$.

4.2 Optimization Objectives and Control Inputs

Experimental results have shown that manipulation of cell-to-cell communication by exposing them to conditioned medium containing elevated AI-2 activity results in increased recombinant protein yield. However, using conditioned medium

(or purified AI-2) would not be a preferred method for accomplishing increased production from a practical point of view. The metabolic pathways indicate that the level of LuxS protein should have an effect on the AI-2 level and the desired recombinant protein (OPH) production induced by IPTG. An objective could be manipulating the level of LuxS protein to maximize AI-2 level and/or OPH production directly. Furthermore, the LuxS level may be controlled dynamically using arabinose to induce LuxS production and manipulating the kinetic rate constant for LuxS production by adjusting the arabinose addition rate. This is an optimal control problem with the rate constant playing the role of control input. Other objectives for optimization, as alternatives or in conjunction with the above objective, could involve a maximization of the speed of increase and of the final level of σ^{32} during a stress response, as well as the sensitivity to the initial distribution of σ^{32} . Another control input would be the σ^{32} targeted antisense RNA.

4.3 Optimal Control for ODE Models

Let us modify the notation of (3.4) to explicitly show a control input $u(t)$

$$\dot{x} = f(x, u) + n(t) \tag{4.1}$$

With the assumption $n(t) = 0$, a deterministic model is given by a set of ordinary differential equations:

$$\dot{x} = f(x, u) \quad x(t_0) = x_0 \tag{4.2}$$

where $u(t)$ subjects to lower and upper bounds:

$$u_L < u(t) < u_U \tag{4.3}$$

The criterion of the performance of the system, i.e. cost function, is defined as a function of the final state:

$$J = \phi(x(t_f)) \quad (4.4)$$

where t_f is the final time of operation. The objective of optimal control is to minimize the cost function J .

Here u is considered to be a piecewise constant function over the time interval $[t_0, t_f]$. The input keeps constant until a switching time is reached, at which point the value changes instantaneously to another constant value, which is held until the next switching time. The control input is assumed to have discontinuities at t_1, t_2, \dots, t_M and can be expressed as (Hasdorff, 1976)

$$u(t) = \sum_{i=0}^M h_i [1(t - t_i) - 1(t - t_{i+1})] \quad t_{M+1} = t_f \quad (4.5)$$

where h_i is the value of the piecewise constant in the time interval $t_i < t < t_{i+1}$ and $1(t)$ is the unit step function. Thus the control input is completely defined by the vector of switching time

$$\tau = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_M \end{bmatrix} \in R^M \quad (4.6)$$

and the vector of input values

$$h = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_M \end{bmatrix} \in R^{M+1} \quad (4.7)$$

The gradient of the cost function J with respect to τ and h is given by (Hasdorff, 1976)

$$g(\tau, h) = \begin{bmatrix} \lambda^T(t_1)[f(x(t_1), h_0) - f(x(t_1), h_1)] \\ \lambda^T(t_2)[f(x(t_2), h_1) - f(x(t_2), h_2)] \\ \vdots \\ \lambda^T(t_M)[f(x(t_M), h_{M-1}) - f(x(t_M), h_M)] \end{bmatrix} \quad (4.8)$$

$$\begin{bmatrix} \int_{t_0}^{t_1} f_u^T(x, h_0) \lambda(t) dt \\ \int_{t_1}^{t_2} f_u^T(x, h_1) \lambda(t) dt \\ \vdots \\ \int_{t_M}^{t_f} f_u^T(x, h_M) \lambda(t) dt \end{bmatrix}$$

where λ is the adjoint vector and satisfies

$$\dot{\lambda} = -f_x^T(x, u) \lambda(t) \quad \lambda(t_f) = \nabla_x \phi(x(t_f)) \quad (4.9)$$

f_x and f_u are given by

$$(f_x)_{ij} = \frac{\partial f_i}{\partial x_j} \quad i, j = 1, \dots, N \quad (4.10)$$

$$(f_u)_i = \frac{\partial f_i}{\partial u} \quad i = 1, \dots, N \quad (4.11)$$

The computation procedure of the cost function and its gradient at a given input $u(t)$ may be outlined briefly as the following:

1. Integrate the system equation (2.1), using the given u , forward in time from t_0 to t_f to get $x(t_0 \sim t_f)$.
2. Having $x(t_f)$, evaluate $\phi(x(t_f))$ and integrate the adjoint system (4.9) backwards in time from t_f to t_0 to get $\lambda(t_0 \sim t_f)$.
3. Use (4.8) to calculate the gradient.

Then the cost function and its gradient are submitted to NLP software to find the optimal solution u^* which minimizes J .

4.4 Optimization for Recombinant Protein Production on A Deterministic Model and Evaluation Based on Stochastic Simulations

From the SPN in Fig. 2.6 , we can write down the deterministic model of σ^{32} genetic circuit for recombinant protein expression as a set of differential equations:

$$\begin{aligned}
\dot{x}_1 &= k_{13}G_1 - k_9x_1 \\
\dot{x}_2 &= k_{14}x_1 + k_{12}x_7 - k_{10}x_2x_8 + k_6G_4x_3 - k_7x_2 + k_8x_3 + k_4G_2x_3 + k_2G_3x_3 \\
\dot{x}_3 &= k_7x_2 - k_6G_4x_3 - k_2G_3x_3 - k_4G_2x_3 - k_8x_3 \\
\dot{x}_4 &= k_4G_2x_3 - k_3x_4 \\
\dot{x}_5 &= k_2G_3x_3 - k_1x_5 \\
\dot{x}_6 &= k_{17}G_5 - k_{18}x_6 - k_{15}x_6x_8 + k_{16}x_9 - k_{19}x_6x_4 \\
\dot{x}_7 &= k_{10}x_2x_8 - k_{12}x_7 - k_{11}x_5x_7 \\
\dot{x}_8 &= k_6G_4x_3 + k_{12}x_7 - k_{10}x_2x_8 - k_5x_8 + k_{11}x_5x_7 + k_{16}x_9 - k_{15}x_6x_8 \\
\dot{x}_9 &= k_{15}x_6x_8 - k_{16}x_9 \\
\dot{x}_{10} &= k_{19}x_6x_4 - k_{20}x_{10}
\end{aligned} \tag{4.12}$$

where the components are listed in Table 2.4, the rate constants are listed in Table 2.1, the genes (they are constants in the model) are listed in Table 2.5.

In this project, σ^{32} translation rate k_{14} is used as the input of optimal control. The effects of σ^{32} antisense on yield of active OPH were studied by Srivastava et al. (2000). In their work, antisense RNA was used to downregulate the σ^{32} mediated response. Although total protein production is lower when the antisense is expressed, the higher specific activity in cultures results in a larger amount of biologically active recombinant protein. Since it is possible to experimentally initiate

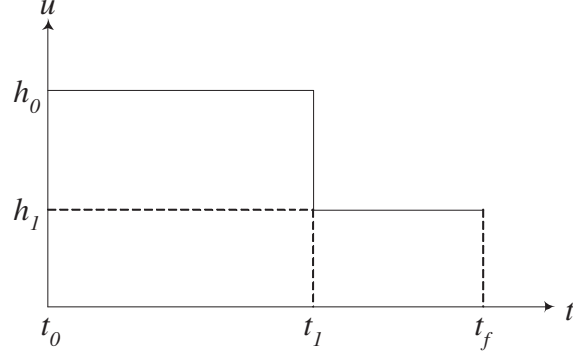


Figure 4.1: Piecewise control input for OPH production

the σ^{32} antisense expression after the protein expression by using different plasmids or different promoters on the same plasmid, the σ^{32} translation rate could be used to as input to maximize OPH production.

Since we use the induction of σ^{32} antisense to adjust the translation rate, the input is considered to be a piecewise constant function with one switch time, as shown in Fig. 4.1. The following values are fixed

$$\begin{aligned} t_0 &= 0 \\ t_f &= 30\text{min} \\ h_0 &= 0.14s^{-1} \end{aligned} \tag{4.13}$$

Thus we have two input variables

$$\tau = t_1 \tag{4.14}$$

$$h = h_1 \tag{4.15}$$

The bounds on input

$$\begin{aligned} 0 &\leq t_1 \leq t_f \\ 0.01s^{-1} &\leq h_1 \leq h_0 \end{aligned} \tag{4.16}$$

The objective of optimal control is to maximize the active OPH level at a fixed final time t_f . The objective function is defined as

$$J = \phi(x(t_f)) = -x_{10}(t_f) \quad (4.17)$$

Its gradient with respect to x

$$\nabla_x \phi = [0, 0, 0, 0, 0, 0, 0, 0, 0, -1] \quad (4.18)$$

According to (4.8), the gradient of the objective function with respect to t_1 and h_1 is given by

$$g(t_1, h_1) = \begin{bmatrix} \lambda^T(t_1)[f(x(t_1), h_0) - f(x(t_1), h_1)] \\ \int_{t_1}^{t_f} f_u^T(x, h_1) \lambda(t) dt \end{bmatrix} \quad (4.19)$$

From the system model (4.12), we can get:

$$f(x(t_1), h_0) - f(x(t_1), h_1) = [0, x_1(t_1)(h_0 - h_1), 0, 0, 0, 0, 0, 0, 0, 0]^T \quad (4.20)$$

$$f_u = [0, x_1, 0, 0, 0, 0, 0, 0, 0, 0]^T \quad (4.21)$$

Substitute (4.20) and (4.21) into (4.19),

$$\begin{aligned} g(t_1, h_1) &= \begin{bmatrix} \lambda_2(t_1)x_1(t_1)(h_0 - h_1) \\ \int_{t_f}^{t_1} -x_1 \lambda_2 dt \end{bmatrix} \\ &= \begin{bmatrix} \lambda_2(t_1)x_1(t_1)(h_0 - h_1) \\ \beta(t_1) \end{bmatrix} \end{aligned} \quad (4.22)$$

where we have defined

$$\beta(t) = \int_{t_f}^t -x_1 \lambda_2 d\tau \quad (4.23)$$

The computation procedure of the value of objective function and its gradient is summarized as the following:

1. Integrate the differential equation (4.12), using the given u , forward in time from t_0 to t_f to get $x(t_0 \sim t_f)$.
2. Having $x(t_f)$, evaluate $\phi(x(t_f))$ and integrate the following differential equations backwards in time from t_f to t_1 to get $\lambda(t_1)$ and $\beta(t_1)$.

$$\begin{aligned}\dot{\lambda} &= -f_x^T \lambda & \lambda(t_f) &= \nabla_x \phi(x(t_f)) \\ \dot{\beta} &= -x_1 \lambda_2 & \beta(t_f) &= 0\end{aligned}\tag{4.24}$$

3. Use (4.22) to calculate the gradient.

The optimization results are

$$\begin{aligned}t_1 &= 10\text{min} \\ h_1 &= 0.01s^{-1}\end{aligned}\tag{4.25}$$

Compared with the yield with constant input (without antisense), the final active protein level increased by 67% with the optimal input. The dynamics of state variables with optimal piecewise input are shown in Fig. 4.2. We can see that the increase of active protein slows down gradually in the first 10 minutes; at the time of 10 minutes, there is a sharp increase of the active protein level, which slows down gradually until the final time. The probability distributions of components across a cell population are shown in Fig. 4.3. The distributions suggest species with lower molecule numbers have relatively larger variances.

By fixing $h_1 = 0.01s^{-1}$ and selecting a range of values for the other input variable, t_1 , between 0 and $t_f = 30\text{min}$, we can simulate the model to obtain the active protein level as a function of t_1 . Those results are shown in Fig. 4.4. We see that the maximum on the curve of the mean value (solid line) seems to be at a

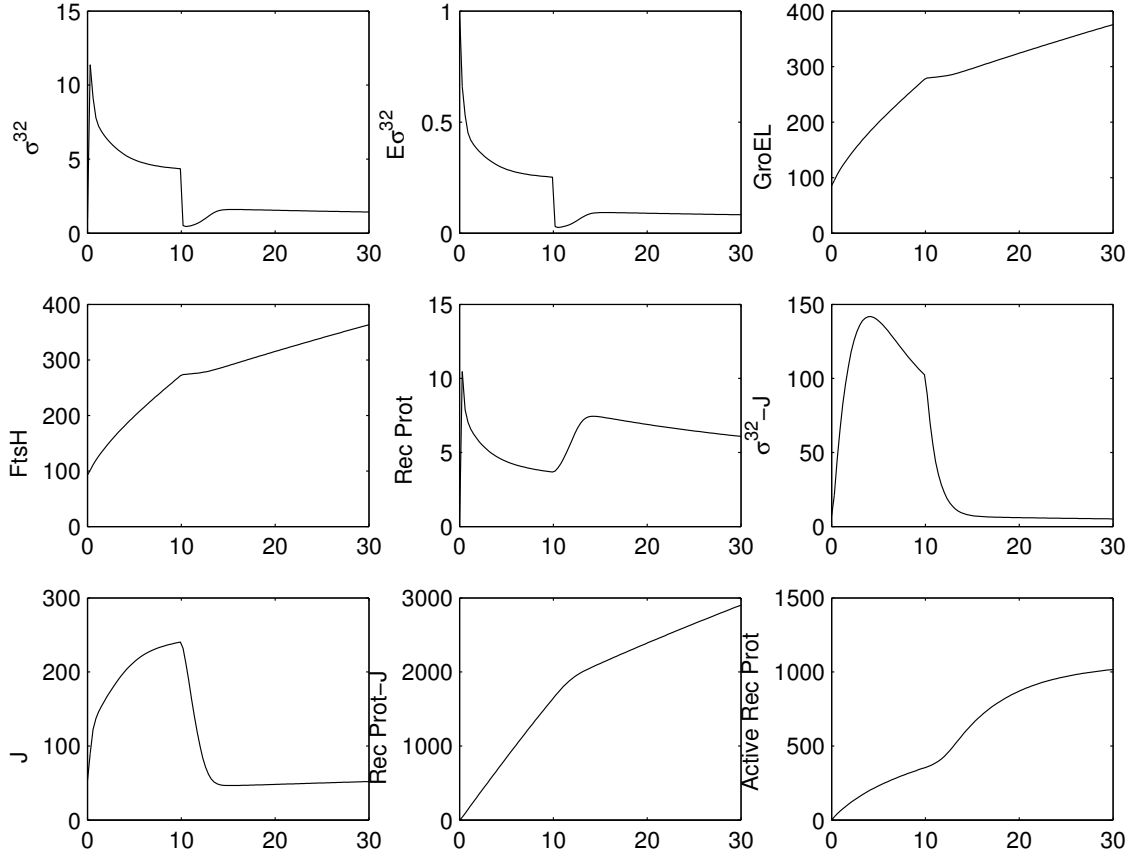


Figure 4.2: Dynamics of state variables in recombinant protein expression with optimal piecewise input: σ^{32} translation rate

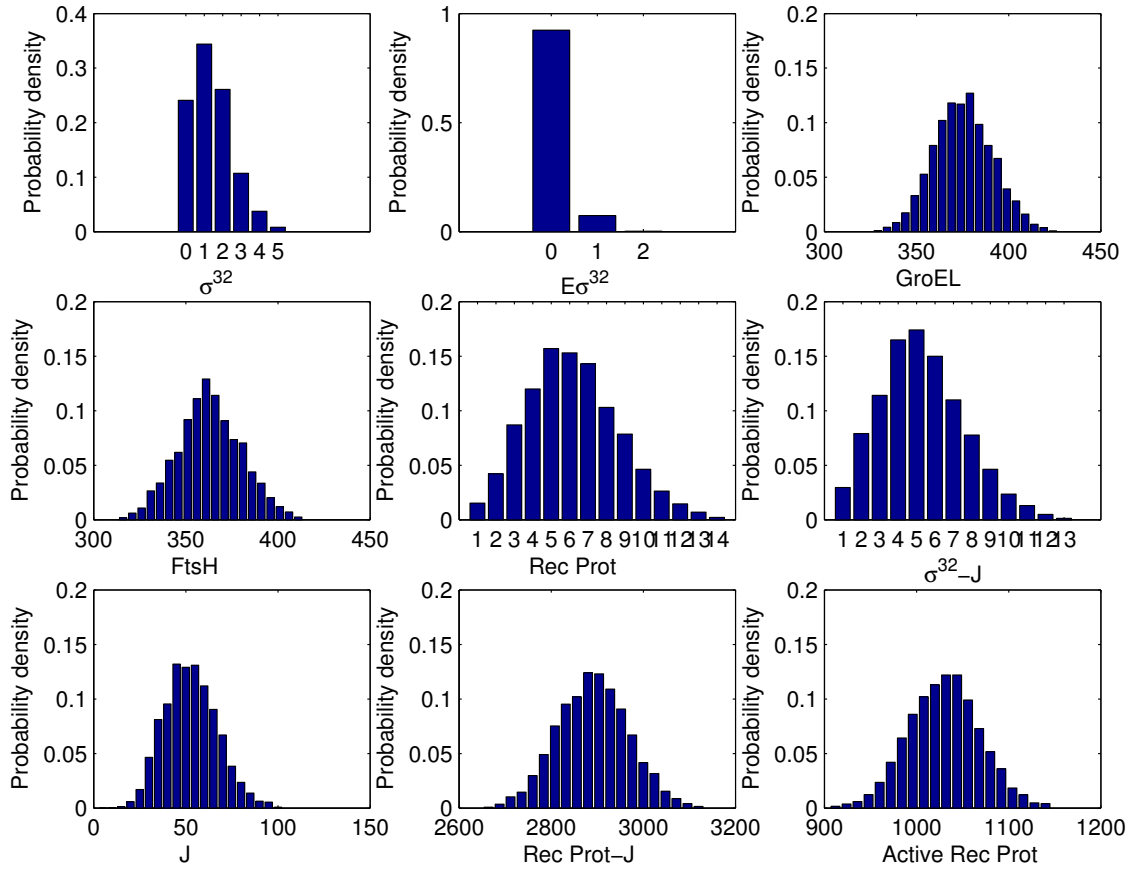


Figure 4.3: Probability distributions of components at 30 min in recombinant protein expression with optimal piecewise input

value for t_1 a little smaller than the one that is optimal for the ODEs. One should keep in mind that SPN simulations with Möbius yield the mean but just like for the variance calculation, there is an error margin that in our simulations has been set to up to 10% for the desired 95% confidence. Usually the error margin turns out to be smaller, about 1-2%, but here the relative flatness of the active OPH level as a function of t_1 results in a difference in the seemingly optimal t_1 . However, this is not important in terms of the result, because the resulting curve is rather flat for times t_1 up to about 15 minutes. SPN simulations also provide the variance that one can expect. This stochastic uncertainty further suggests that the active protein production will not deviate much from the expected optimum for such times t_1 . This result suggests that there is substantial tolerance in the amount of time we can have between the recombinant protein induction and the antisense induction.

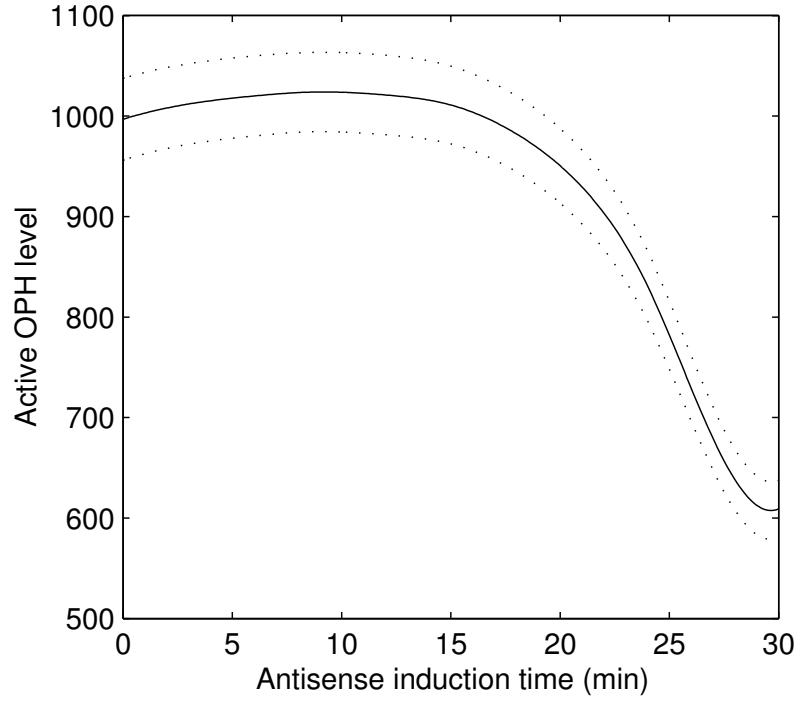


Figure 4.4: Solid line: The final active OPH level (from SPN simulations) as a function of t_1 , i.e. the time of antisense induction. Surrounding dotted lines: \pm standard deviation.

Chapter 5

Conclusions and Future Work

When *E. coli* cells are subject to stress conditions, such as ethanol shock or expression of recombinant protein, the transcription factor σ^{32} initiates responses that increase the production of heat shock proteins and repair denatured proteins. In the work of Srivastava et al. (2000), antisense RNA was used to downregulate σ^{32} -mediated stress responses. While the total yield of recombinant protein is lower when antisense is expressed, the higher specific activity indicates a larger amount of biologically active protein.

In genetic regulatory circuits, many molecules are present in very low concentrations. As a result, the discrete nature becomes important and the stochastic effects may invalidate the deterministic equations. A more broadly applicable approach to model such systems is to use chemical master equations, which describe the reaction network as a continuous time, discrete state-space and stochastic process. CME are impossible to solve for all but the simplest systems, but they can be simulated numerically. Stochastic Petri Nets simulation is equivalent to the First Reaction Method proposed by Gillespie (1976) for numerical simulation of CME. Based on previous work (Srivastava et al., 2001), we further developed the SPN models for ethanol stress response and recombinant protein production. The model parameters were further tuned and the model was extended to include the produc-

tion of biologically active recombinant protein. The simulation results showed good match to experimental data reported in Srivastava et al. (2000).

The optimal control for genetic circuits in this work aims at developing control policies for dynamically optimizing pathway performance. However, using SPN directly as part of an optimization problem results in tremendous computational difficulties. A possible alternative approach is to use Langevin-type models which are more suitable for use in an optimal control problem. The system is described as

$$\dot{x} = f(x) + n(t) \quad (5.1)$$

where $f(x)$ is the deterministic kinetics of the reaction network, $n(t)$ is a vector of noise variables. $n(t)$ represents the stochasticity of the system and its covariance matrix Q can be determined from the distribution information of x obtained from SPN simulations. A Langevin model was developed for ethanol stress response. At first a diagonal Q was assumed and its elements are obtained from a constrained optimization. Then a non-diagonal element of Q was included to reduce the error in matching diagonal elements.

As the first attempt, optimal control policies were developed for the deterministic part of a Langevin model. Since it is possible to experimentally initiate the σ^{32} antisense expression after the recombinant protein expression by using different plasmids or different promoters on the same plasmid, the timing of antisense induction can be used to maximize the active protein production. The change in σ^{32} translation rate was used to simulate the effects of antisense and was considered to be a piecewise control input. Optimization results showed an increase of active

protein level by 67% with optimal input. The robustness of the results with respect to uncertainties in the model was also discussed.

Future work may include:

1. Further improving the Langevin models to reduce the error in the match of variance information. Utilize Langevin models for robust control by interpreting the noise terms as “uncertainty” in the models (Morari and Zafiriou, 1989).
2. Develop sensitivity analysis formalism for stochastic models (Feng et al., 2004). Study the robustness of genetic circuits with respect to model parameters and other environmental factors (El-Samad et al., 2002).
3. Combine the σ^{32} circuit with “quorum sensing” (cell to cell communication) circuit as well as study the interplay between these circuits.

BIBLIOGRAPHY

- [1] Arkin, A., J. Ross, and H. H. McAdams (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149: 1633-1648.
- [2] Arnold, L. (1974) *Stochastic differential equations: theory and applications*. John Wiley & Sons, New York.
- [3] Arsene, F., T. Tomoyasu, and B. Bukau (2000) The heat shock response of *Escherichia coli*. *International Journal of Food Microbiology* 55, 3-9.
- [4] Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* 10, 411-421.
- [5] Birge, J. R. and F. Louveaux (1997) *Introduction to Stochastic Programming*, Springer-Verlag.
- [6] Blaszcak, A., M. Zylicz, C. Georgopoulos, and K. Liberek (1995) Both ambient temperature and the DnaK chaperone machine modulate the heat shock response in *Escherichia coli* by regulating the switch between σ^{70} and σ^{32} factors assembled with RNA polymerase. *The EMBO J.* 14: 5085-5093.
- [7] Braun, D., S. Basu, and R. Weiss (2005) Parameter estimation for two synthetic gene networks: a case study. *IEEE International Conference on Acoustics, Speech, and Signal Processing*
- [8] Brendel, V., and A. S. Perelson (1993), *J. Mol. Biol.*, 229, 860-872.
- [9] Brillouin, L. (1950), *Phys. Rev.*, 78, p. 627.
- [10] Casinovi, G. (2002) An algorithm for frequency-domain noise analysis in non-linear systems (2002) *Design Automation Conference*, New Orleans, Louisiana, USA June 2002.
- [11] Cheng, J.-H. (2001) *Robust Model-Based Steady-State Feedback Optimization for Chemical Plants*, Ph.D. Dissertation, Univ. of Maryland.
- [12] Cheng, J.-H. and E. Zafiriou (2002) Results Analysis for Iterative Feedback Steady State Optimization, accepted for inclusion in *Proc. Int. Fed. Automatic Control World Congress*, Barcelona, Spain, July 2002.
- [13] Cox C. D., G. D. Peterson, M. S. Allen, J. M. Lancaster, J. M. McCollum, D. Austin, L. Yan, G. S. Sayler and M. L. Simpson (2003) Analysis of Noise in Quorum Sensing. *OMICS* 7, 317-334.
- [14] Elf, J. and M. Ehrenberg (2003) Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Research* 13, 2475-2484.

- [15] El-Samad, H., M. Khammash, H. Kurata, and J. Doyle (2002) Robustness Analysis of the Heat Shock Response in *E. coli*, *Proceedings of the American Control Conference*, 2002, 1742-1747.
- [16] Érdi, P., and J. Tóth. (1988) *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Princeton University Press, Princeton, N.J.
- [17] Feng, X., S. Hooshangi, D. Chen, G. Li, R. Weiss, and H. Rabitz (2004) Optimizing genetic circuits by global sensitivity analysis. *Biophysical Journal* 87, 2195-2202.
- [18] Gamer, J., H. Bujard, and B. Bukau (1992) Physical interaction between heat shock proteins DnaK, DnaJ, and GrpE and the bacterial heat shock transcription factor σ^{32} . *Cell* 69: 833-842.
- [19] Gamer, J., G. Multhaup, T. Tomoyasu, J. S. McCarty, S. Rudiger, H. Schonfeld, C. Schirra, H. Bujard, and B. Bukau (1996) A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the *Escherichia coli* heat shock transcription factor σ^{32} . *EMBO J.* 15: 607-617.
- [20] Gibson, M. and E. Mjolsness (2001), "Modeling the Activity of Single Genes," in *Computational Modeling of Genetic and Biochemical Networks*, J. M. Bower and H. Bolouri (Eds.), MIT Press, Cambridge, MA, pp. 1-48.
- [21] Gillespie, D.T. (1976) General method for numerically simulating stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* 22, 403-434.
- [22] Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* 81, 2340-2361.
- [23] Gillespie, D.T. (1992) A rigorous derivation of the chemical master equation. *Physica A* 188, 404-425
- [24] Gillespie, D.T. (2000) The Chemical Langevin equation. *Journal of Chemical Physics* 113, 297-306
- [25] Goss, P. J., and J. Peccoud (1998) Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. U S A* 95: 6750-5.
- [26] Gottesman, S. (1996) Protease and their targets in *Escherichia Coli*.. *Annual Review of Genetics* 30, 465-506.
- [27] Gunawan, R., Y. Cao, L. Petzold, and F. J. Joyle III (2005) Sensitivity analysis of discrete stochastic systems. *Biophysical Journal* 88, 2530-2540.

- [28] Hasdorff, L. (1976) *Gradient optimization and nonlinear control*. John Wiley & Sons, New York.
- [29] Herendeen, S. L., R. A. Vanbogelen, and F. C. Neidhardt (1979) Levels of major proteins of *Escherichia coli* during growth at different temperatures. *Journal of Bacteriology* 139, 185-194.
- [30] Jong H. D. (2002) Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology* 9, 67-103.
- [31] Kall, P. and S. W. Wallace (1995), *Stochastic Programming*, John Wiley and Sons.
- [32] Kanemori, M., H. Mori, and T. Yura (1994) Induction of heat shock proteins by abnormal proteins results from stabilization and not increased synthesis of σ^{32} in *Escherichia coli*. *J. Bacteriol.* 176, 5648-5653.
- [33] Kepler, T. B. and T. C. Elston (2001) Stochasticity in Transcriptional regulation: origins, consequences, and mathematical representations. *Biophysical Journal* 81, 3116-3136.
- [34] Kurata H., H. El-Samad, T. M. Yi, M. Khammash, J. C. Doyle (2001) Feedback Regulation of the Heat Shock Response in *E. coli*, *Proceedings of the 40th IEEE Conference on Decision and Control*, 837-842.
- [35] Lamb, T. D. (1996) Gain and kinetics of activation in the G-protein cascade of phototransduction. *Proc. Natl. Acad. Sci. USA* 93, 566-570.
- [36] Lesley, S. A., N. E. Thompson, and R. R. Burgess (1987) Studies of the role of the *Escherichia coli* heat shock regulatory protein σ^{32} by the use of monoclonal antibodies. *J. Biol. Chem.* 262, 5404-5407.
- [37] McAdams, H. H., and A. Arkin (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U S A* 94, 814-819.
- [38] McFee (1971) Self-Rectification in Diodes and the Second Law of Thermodynamics, *textitAmer. J. Physics*, 39, p. 814.
- [39] Miletic, I. P. and T. E. Marlin (1998) On-line Statistical Results Analysis in Real-Time Operations Optimization, *Ind. Eng. Chem. Res.*, 37, 3670-3684.
- [40] Morari, M. and E. Zafiriou (1989) *Robust Process Control*. Prentice-Hall, Englewood Cliffs, NJ.
- [41] Muffler, A., M. Barth, C. Marschall, and R. Hengge-Aronis (1997) Heat shock regulation of σ^S turnover: a role for DnaK and relationship between stress responses mediated by σ^S and σ^{32} *Escherichia coli*. *J. Bacteriol.* 179, 445-452.
- [42] Murray, J. A. H. (1992) *Antisense RNA and DNA*. Wiley-Liss, New York.

- [43] Neidhardt, F. C., R. A. VanBogelen, and V. Vaughn (1984) The genetics and regulation of heat-shock proteins. *Ann. Rev. Genet.* 18, 295-329.
- [44] Neidhardt, F. C., J. L. Ingraham and M. Schaechter (1990) *Physiology of the bacterial cell : a molecular approach*. Sunderland, Mass. : Sinauer Associates.
- [45] Neidhardt, F. C., R. Curtis, III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magansik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umberger (1996) *Escherichia coli and Salmonella: Cellular and Molecular Biology*. ASM, Washington D.C.
- [46] Ozbudak, E. M., M. Thattai, I. Kurster, A. D. Grossman, and A. v. Oudenaarden (2002) Regulation of noise in the expression of a single. *Nature Genetics* 31, 69-73.
- [47] Paulsson, J. (2004) Summing up the noise in gene networks. *Nature* 427, 415-418.
- [48] Peccoud, J. and B. Ycart (1995) Markovian modeling of gene product synthesis. *Theoretical Population Biology* 48, 222-234.
- [49] Pedersen, S., P. L. Bloch, S. Reeh, and F. C. Neidhardt (1978) Patterns of protein synthesis in E. coli: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* 14, 179-190.
- [50] Rao, C. V., D. M. Wolf, and A. P. Arkin (2002) Control, exploitation and tolerance of intracellular noise. *Nature* 420, 231-237.
- [51] Risken, H. (1989) *The Fokker-Planck Equation*, Springer-Verlag.
- [52] Sanders, W. H. (1995) *UltraSAN User's Manual*. Center for Reliable and High-Performance Computing, University of Illinois at Urbana-Champaign.
- [53] Saucedo, V. M., and M. N. Karim (1998) Real Time Optimal Feeding in a Fermentor Using a Markov Decision Algorithm, *Computers Chem. Engng.*, 22, 547-558.
- [54] Simpson, M. L., C. D. Cox, and G. S. Sayler (2003) Frequency domain analysis of noise in autoregulated gene circuits. *PNAS* 100, 4551-4556.
- [55] Srivastava, R. (1999), *Analysis and modeling of the effects of mRNA antisense targeted against the σ^{32} stress response sigma factor in Escherichia coli*, Ph.D. Dissertation, Univ. of Maryland.
- [56] Srivastava, R., Cha, H.J., Peterson, M.S. and Bentley, W.E. (2000) Antisense downregulation of σ^{32} as a transient metabolic controller in *Escherichia coli*: effects on yield of active organophosphorus hydrolase. *Appl Environ Microbiol* 66, 4366-71.

- [57] Srivastava, R., Peterson, M.S., and W.E. Bentley (2001) Stochastic kinetic analysis of the *Escherichia. coli* stress circuit using σ^{32} -targeted antisense, *Biotechnol. Bioeng.* 75,120-129.
- [58] Strauss, D. B., W. A. Walter, and C. A. Gross (1987) The heat shock response of *E. coli* is regulated by changes in the concentration of σ^{32} . *Nature* 329, 348-391.
- [59] Strauss, D. B., W. A. Walter, and C. A. Gross (1989) The activity of σ^{32} is reduced under conditions of excess heat shock protein production in *Escherichia coli*. *Genes Develop.* 3, 2003-2010.
- [60] Thattai, M., and A. v. Oudenaarden (2001) Intrinsic noise in gene regulatory networks. *PNAS* 98, 8614-8619.
- [61] Tomoyasu T., Gamer J., Bukau B., Kanemori M., Mori H., Rutman A. J., Oppenheim A. B., Yura T., Yamanaka K., Niki H., Hiraga S. and Ogura T. (1995) *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor σ^{32} . *EMBO J.* 14, 2551-2560.
- [62] Van Kampen, N. G. (1992) *Stochastic Processes in Physics and Chemistry*, North Holland.
- [63] Weiss, R., S. Basu, S. Hooshangi, A. Kalmbach, D. Karig, R. Mehreja, and I. Metravali (2003) Genetic circuit building blocks for cellular computation, communications, and signal processing. *Natural Computing* 2, 47-84.
- [64] Yokobayashi, Y., R. Weiss, and F. H. Arnold (2002) Directed evolution of a genetic circuit. *PNAS* 99, 16587-16591.
- [65] Zheng, Q., and E. Zafiriou (1995) Nonlinear System Identification for Control Using Volterra-Laguerre Expansion, *Proc. Amer. Control Conf.*, Seattle, WA, 2195-2199.
- [66] Zhou, Y., N. Kusukawa, J. W. Erickson, C. A. Gross, and T. Yura (1988) Isolation and characterization of *Escherichia coli* mutants that lack the heat shock sigma factor σ^{32} . *Journal of Bacteriology* 170, 3640-3649.